Syrian Arab Republic
Ministry of Higher Education
Syrian Virtual University
BIS ماجستير التأهيل والتخصص في المعلوماتية الحيوية

Designing a virtual multi-epitopes vaccine against *Vibrio cholerae* using immunoinformatics

Presented by: Pharmacist Rafa Saleh / rafa_157196

Supervised by: Dr. Abdul Qader Abbady

Table of contents:

1- Introduction:

Cholera is an acute watery diarrheal infection that can lead to death if left untreated, it is caused by ingestion of food or water contaminated with the bacterium *Vibrio cholerae* (*V. cholerae*) [1]. *V. cholerae* is a gram-negative, highly motile curved rods with a single polar flagellum. It is classified by the composition of its major surface antigen (O) from lipopolysaccharide into nearly 206 serogroups. But only two serogroups of *V. cholerae* (O1 and O139) are considered as causative agents of epidemic cholera [2]. Currently there are three World Health Organization (WHO) pre-qualified oral cholera vaccines (OCV): Dukoral®, Shanchol™, and Euvichol® [3]. OCVs are not recommended in infants. Furthermore, limited protection was conferred by OCVs among children less than 5 years of age. Over and above that, two or three repeated (14-day intervals between the two doses) vaccination doses are necessary for longer protection in cholera endemic situations [4]. Some limitations may be inherent to the formalin and heat used to kill the *V. cholerae* strains in these vaccines, which may destroy or alter protein epitopes [5] which could lead to lack of immunogenicity. Thus, it was suggested to develop a new generation of vaccines using immunoinformatics, the most informative and advantageous device for vaccine design [6].

Immunoinformatics as a subset of bioinformatics is a new approach that facilitates analyzing the enormous amounts of immunologic data obtained from experimental research using a variety of tools and databases [7]. Reverse vaccinology (RV) is a new immunoinformatics concept. The original idea behind RV was to start in-silico to screen the entire genome of a pathogen to identify genes that encode proteins with the attributes of good vaccine targets [8]. One of the main applications of immunoinformatics is developing B and T cell epitope prediction algorithms which decrease the time and costs required for experimental analysis [7]. The epitopes are selected on the basis that they were accessible to immune system surveillance using some form of informatic-based prediction methodology or a set of methodologies [9]. The use of immunoinformatics has been accelerated toward the design of multi epitope vaccines. [10]. There are a lot of databases of B-cell and T-cell epitopes. Also, there are a lot of web servers and tools for prediction of B and T cell epitopes.

By immunoinformatics, multi-epitope subunit vaccine against HIV infection was designed by the combination of Cytotoxic T-lymphocyte cells (CTL), Helper T-lymphocyte cell (HTL) and B lymphocyte cells (BCL) epitopes along with suitable adjuvant and linkers [11]. Also, four antigenic proteins of Leishmania donovani were chosen and their T-cell and B-cell epitopes were identified, utilizing them for in-silico chimeric vaccine designing [12]. Likewise, a novel multi-epitope vaccine was designed to induce cellular, humoral, and innate immune responses against

Staphylococcus aureus using immunoinformatics tools[13]. As another example, an epitope vaccine against cholera was designed to overcome the low immunogenicity [14].

2- Dataset:

V. cholerae O1 biovar El tor str. N16961 was selected as a reference strain in this study. According to our survey in the medical literature, the most supported proteins which frequently mentioned as reasonable antigens to include in a cholera vaccine formulation were nominated. Their Fasta formats were retrieved from Uniprot to be subjected to antigenicity, virulence and subcellular localization tests through suitable tools and web servers. Actin (ACTB), the housekeeping protein in the homo-sapiens has no immunogenic significance, its Fasta format was retrieved from Uniprot to be subjected to the same tools and web servers in order to use Actin as a negative control. Table (1) below describes the chosen proteins from medical literature.

Table (1) Shows the chosen proteins from medical literature with their references and accession number in Uniprot:

	Protein	Reference	Accession Number in Uniprot
1	Toxin coregulated pilin A TcpA	[16]	Q60153
2	Toxin coregulated pilin F TcpF	[17]	P0C6Q5
3	Outer membrane protein U OmpU	[18] [19] [20]	P0C6Q6
4	Outer membrane protein W OmpW	[18] [20]	P17266
5	Porin OmpT	[18] [20]	Q9KQZ3
6	Accessory colonization factor A AcfA	[18]	H9L4S5
7	Lipoprotein NlpD	[18]	Q9KUI7
8	Outer membrane protein TolC	[20]	Q9K2Y1

9	FlgO domain-containing protein FlgO	[20]	Q9KQ00
10	Flagellar biosynthesis protein FlgP	[20]	Q9KQ01
11	flagellin A FlaA	[21]	P0C6C3
12	Actin (Negative Control) ACTB	_	P60709

The aforementioned proteins were checked by the VFDB tool. VFDB tool is an integrated and comprehensive online resource for curating information about virulence factors of bacterial pathogens [22] [23]. Through VFDB, each protein was subjected to BLAST sequence-similarity search using Blastp, they were compared with the Virulence Factor core dataset which contains proteins associated with experimentally verified virulence factors. Also each protein was checked by Vaxijen 2.0 server [24] in order to evaluate the capacity of these epitopes to prompt an immune response. Threshold for this model was 0.5 [24]. Subcellular localization was checked using psortb, the most precise bacterial protein subcellular localization (SCL) predictor since it was first made available in 2003 [25]. Periplasmic, outer membrane and extracellular proteins are exposed at the surface and they are more accessible to the immune system. Results are shown in Table (2) below.

Table (2) shows VFDB, Vaxijen and psortb predictions:

No.	Protein	VFDB	Vaxijen		Subcellular Localization
		Identities	Overall Prediction for the Protective Antigen	Probable	
1	TcpA	95%	0.3670	Non-Antigen	Extracellular
2	TcpF	99%	0.6308	Antigen	Unknown
3	OmpU	68%	0.7413	Antigen	Outer Membrane
4	OmpW	No hits found	0.7778	Antigen	Outer Membrane
5	OmpT	22%	0.7438	Antigen	Outer Membrane

6	AcfA	96%	0.7646	Antigen	Outer Membrane
7	NlpD	No hits found	0.7922	Antigen	Outer Membrane
8	TolC	20%	0.5453	Antigen	Outer Membrane
9	FlgO	100%	0.2860	Non-Antigen	Unknown
10	FlgP	100%	0.6704	Antigen	Unknown
11	FlaA	100%	0.6734	Antigen	Extracellular
12	ACTB (negative control)	No hits found	0.3510	Non-Antigen	Cytoplasmic

The cholera disease is caused by enteropathogenic organism *V. cholerae* through intestinal colonization and elaboration of a potent enterotoxin known as cholera toxin (CTX) [26]. CTX is the major contributing factor for profuse diarrhea, it is produced by the epidemic causing strains of *V. cholerae* (O1 or O139 serogroups). It should be noted here that there are also strains of O1 and O139 serogroup which do not produce CTX, and are not involved in epidemics. Conversely, there are occasional strains of serogroups other than O1 or O139 that are clearly pathogenic, either by the production of CTX or other virulence factors [27]. Lönnroth and Holmgren and others demonstrated that CTX is made up of two types of subunits [15]. Five B subunit (CTB) which binds holotoxin to the cell receptor and one A subunit (CTA) which provides intracellular activity [27]. The secretory IgA class of antibodies act through inhibition of intestinal attachment and subsequent colonization of vibrio. Antibodies to *V. cholerae* Lipopolysaccharide (LPS) mediate protection against cholera. But also the existence of non-LPS protective antigens has also been documented to play an important role in protection via inhibition of intestinal colonization of vibrios [26]. Tcp is thought to be a polymer composed of a single structural subunit that facilitates microcolony formation on the epithelial cell surface [28]. TcpF may be a reasonable antigen to include in a polyvalent subunit vaccine formulation [17]. The TcpA may represent an important new immunogen for incorporation into improved vaccines [29]. The porin proteins may be considered major protective antigens of V. cholera. The outer membrane proteins (Omp), OmpU and OmpT contribute to *V. cholerae* virulence. These porins were suspected to be involved in virulence because their expression is regulated by ToxR, which also regulates CTX and Tcp. ToxR activates the transcription of OmpU and represses the transcription of OmpT. OmpU is more protective (compared to OmpT) against the bactericidal effects of bile salts and other anionic detergents [28]. OmpW is thought to be very immunogenic and may correspond to one of the major immunogenic proteins [30]. TolC is a major outer membrane protein involved in bacterial multidrug resistance and survival of pathogens during infection in several gram-negative bacteria.

There is an in vivo role of TolC in bile resistance and colonization [32]. The outer membrane proteins FlgO and FlgP are also playing a role in immune response generation [20]. Accessory colonization factors (Acfs) have shown to be important for colonization of the small intestine [28]. In addition, the role of lipoprotein NlpD, has been studied in reference to cell division and intestinal colonization by the pathogen. As septal peptidoglycan amidase (AmiB) is involved in separation of daughter cells at the end of cell division process and AmiB is regulated by NlpD in *V. cholerae* [18]. FlaA protein is already known in literature as a vaccine candidate [21]. FlaA is essential for assembly and function of the flagellum. Both expression of flagella and motility contribute to colonization and virulence [33].

3- Methodology:

1- Proteins' sequences retrieved in Fasta format.
2- Epitope prediction using bioinformatics tools (considering the classification of Human Leukocyte Antigen (HLA) Supertypes) for proteins.

- Cytotoxic T Lymphocyte (CTL) and Helper T Lymphocyte (HTL) epitope prediction.
- B Lymphocyte epitope prediction.

3- Population coverage calculation.
4- Vaccine construction.
5- Allergenicity, antigenicity, solubility and physicochemical aspects predictions for the designed vaccine.
6- 2-dimensional structure analysis.
7- 3-dimensional structure modeling and refinement.
8- Discontinuous (conformational) B-cell epitopes prediction.
9- Immune response simulation analysis.
10- Molecular docking.
11- Codon optimization and cloning.

3-1- Proteins' sequences retrieval in Fasta format from Uniprot.

3-2- Epitope prediction:

The human leukocyte antigen (HLA) system (the major histocompatibility complex [MHC] in humans) is an important part of the immune system and is controlled by genes located on chromosome 6. It encodes cell surface molecules specialized to present antigenic peptides to the T-cell receptor (TCR) on T cells [36].
The HLA system consists of two classes: class I region contains HLA-A, HLA-B and HLA-C genes and class II region contains HLA-DR, HLA-DQ and HLA-DP [34].

3-2-1- Cytotoxic T Lymphocyte (CTL) epitope prediction:

CTL epitopes of the proteins were predicted using the Immune Epitope Database Analysis Resource website which provides a collection of tools for the prediction and analysis of immune epitopes. It serves as a companion site to the Immune Epitope Database (IEDB), a manually curated database of experimentally characterized immune epitopes. CTL epitopes were predicted using netmhcpan_el method [37] [38] [39].

It makes sense to divide antigenicity into immunogenicity and allergenicity. Antigenicity and immunogenicity are generally used to describe each antigen. Antigenicity is the ability to specifically combine with the final products of the immune response (i.e., secreted antibodies and/or surface receptors on T cells). Immunogenicity is the ability to induce a humoral and/or cell-mediated immune response. Although all molecules that are immunogenic are also antigenic, the reverse is not true. Allergenicity refers to the ability of an antigen to induce an abnormal immune response, which is an overreaction and different from a normal immune response in that it does not result in a protective/prophylaxis effect but instead causes physiological function disorder or tissue damage [40]. The sequence of each epitope was entered into Vaxijen 2.0 server in order to evaluate the capacity of these epitopes to prompt an immune response. Threshold for this model is 0.5, since most of the models had their highest accuracy at a threshold of 0.5 [24]. Some peptides are more immunogenic than others and therefore more likely to be T-cell epitopes. For that, the accepted epitopes were entered into Class I Immunogenicity [41] from IEDB Analysis Resource in order to compute immunogenicity score. The epitopes which got positive scores were accepted. The epitopes were entered into the AllerTOP v. 2.0 tool which is a bioinformatics tool for allergenicity prediction [42]. The components of the vaccine must not present allergic responses. ToxinPred was used to determine the toxicity of accepted epitopes [43] [44]. ToxinPred is a unique in-silico method of its kind, which was useful in predicting toxicity of peptides/proteins. It was useful in designing least toxic peptides and discovering toxic regions in proteins [44].

3-2-2- Helper T Lymphocyte (HTL) epitope prediction:

HTL epitopes of the proteins were predicted using NN-align [45], a novel artificial neural network-based method that allows for simultaneous identification of the MHC class II binding core and binding affinity; it outperforms other MHC class II prediction methods [46]. Antigenicity, allergenicity and toxicity predictions for epitopes was done for HTL epitopes as explained above in CTL epitopes as well as the IFN-γ prediction for HTL epitopes by IFN-γ epitope server (SVM based method) as macrophages produce large amounts of cytokines such as interferon (IFN)- γ which has a regulatory role in adaptive immune responses [47]. The accepted epitopes were entered into IL4pred which is an In-Silico platform for designing and discovering of Interleukin-4 inducing peptides [48].

3-2-3- Linear B Lymphocyte (LBL) epitope prediction:

B-cell epitopes play a vital role in the development of peptide vaccines. ABCpred server was used to predict linear B cell epitope regions using an artificial neural network. This server assists in locating epitope regions that are useful in selecting synthetic vaccine candidates [49] [50].

3-3- Population Coverage Calculation:

MHC molecules are extremely polymorphic (over a thousand different human MHC (HLA) alleles are known) and the epitopes elicit a response only in individuals that express an MHC molecule capable of binding that particular epitope. In the design of peptide-based vaccines and diagnostics, the issue of population coverage in relation to MHC polymorphism is further complicated by the fact that different HLA types are expressed at dramatically different frequencies in different ethnicities [51]. A web-based tool from IEDB Analysis Resource has been developed to predict population coverage of T-cell epitope-based diagnostics and vaccines based on MHC binding and/or T cell restriction data [51].

3-4- Vaccine construction:

The epitopes which had been determined previously by various immunoinformatics software were connected and linked together with the aid of separate linkers in order to develop the final vaccine. A multi-epitope vaccine without linkers may result in a new protein with unknown properties or may result in the formation of neoepitopes/junctional epitopes [53]. The flexibility or rigidity of the linker between two fused proteins is an important parameter that affects the function of fusion proteins [54]. Flexible linkers are generally composed of small, non-polar or polar residues [56] and using them might increase some biological activities of the protein [55]. Rigid linkers offer efficient separation of the functional domains by keeping a fixed distance with minimal interference between the epitopes thereby maintaining their individual functional properties. This helps in the effective separation of domains in a bifunctional fusion protein [53]. B subunit of cholera toxin (CTB) was added as an adjuvant to increase the immunogenicity of the vaccine [52]. CTB sequence was retrieved from Uniprot (Uniprot accession number: 01556).

Multiple combinations of the vaccine candidate (depending on studies in the medical literature) were proposed and discussed. And the best combination was approved as a final format for the vaccine. Another one was used for comparison.

3-5- Physicochemical and other properties prediction:

The physicochemical properties of the proposed vaccine candidates were assessed by ProtParam tool / Expasy [57]. ProtParam is a tool which allows the computation of various physical and chemical parameters for a given protein. The computed parameters include the molecular weight, theoretical isoelectric point (pI), amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY). Solubility predictions were assisted by both Protein-Sol web tool [58] and SOLpro predictor [59]. Protein-sol is a simple and free, web-based suite of theoretical calculations and predictive algorithms for understanding protein solubility and stability. The scaled solubility value is the predicted solubility. The population average for the experimental dataset is 0.45 [58]. SOLpro predicts the propensity of a protein to be soluble upon overexpression in E. coli using a two-stage SVM architecture based on multiple representations of the primary sequence [59]. The antigenicity, allergenicity, homology with homo-sapiens proteins and toxicity properties of the vaccine candidate was assessed using various tools. The best candidate was chosen over all candidates. Another good candidate was taken for comparison.

3-6- Secondary structure analysis:

The secondary structural features include a-helix, b-strand and random coils that were evaluated using the SOPMA server [62] .

3-7- Tertiary structure modeling and refinement:

The Phyre2 server [63] was used in the homology modeling candidates using intensive modeling mode. Ramachandran plots for the modeled candidates were created by Prochek [64]. The models were refined using GalaxyRefine web server [75]. Before and after refinement models were assessed by Ramachandran plots and also were inserted into ProSA-web protein structure analysis [65] [66]. ProSA is a tool widely used to check 3D models of protein structures for potential errors. The z-score indicates overall model quality and measures the deviation of the total energy of the structure with respect to an energy distribution derived from random conformations. Z-scores outside a range characteristic for native proteins indicate erroneous structures.

3-8- Defining discontinuous B-cell epitopes (conformational):

B-cell epitopes play an important role in humoral response. The refined 3D protein vaccine model was subjected to B-cell epitopes prediction using Ellipro Server [68] in order to predict

conformational B-cell epitopes.

3-9- Immune response simulation:

The probability of the designed vaccine inducing both humoral and cellular immune responses was further assessed using the C-IMMSIM server [67].

3-10- Molecular docking:

Patchdock [70] [71] was used for ranking the top interaction models between the vaccine candidate and receptor. These top models were refined in Firedock [77] [78]. Discovery Studio 2021 program was used to visualize the results.

3-11- Codon adaptation and in-silico cloning:

Java Codon Adaptation Tool (JCat) [80] was used to corroborate codon compliance by optimizing the vaccine candidate sequence. The designed candidate was optimized according to E. coli strain K12, which is known for its quick reproduction rates and ability to survive [69]. Two restriction enzymes NcoI and XhoI were added respectively to the N- and C-terminal of the protein to finally insert it into the pET23d (+) vector using the Geneious software.

4- Results:

4-1- Proteins' sequences retrieval in Fasta format from Uniprot.

4-2- Epitope prediction:

4-2-1- Cytotoxic T Lymphocyte (CTL) epitope prediction:

The most common alleles for HLA class I A locus in the Syrian population were: A*02, A*24, A*01, A*03; for B locus were: B*35, B*51, B*44, B*52 and for C locus were: C*04, C*07, C*12, C*06 [34] and they were used in predicting the epitopes by Immune Epitope Database Analysis Resource MHC class I prediction. The best 5 results for each protein (according to the fact that

high Score = good binder) were subjected to further antigenicity, immunogenicity, allergenicity and toxicity predictions using Vaxijen 2.0 server, Class I Immunogenicity [41] from IEDB Analysis Resource, AllerTOP v. 2.0 and ToxinPred respectively in order to choose the best ones ever. Results are shown below in the tables (3-13)

Table (3) shows Cytotoxic T Lymphocyte (CTL) epitopes prediction for TcpA protein with their scores and antigenicity, immunogenicity, allergenicity and toxicity predictions for each epitope:

Protein1_TcpA						
Epitope	Score	Prediction for the Protective Antigen	Probable	Immunogenicity score	Allergenicity	Toxicity
TLLEVIIVL	0.939595	0.4338	Non-Antigen	0.39907	Non-Allergen	Non-Toxin
QLANGLVSL	0.884819	-0.0927	Non-Antigen	-0.03307	Allergen	Non-Toxin
QLFKKKFVK	0.868859	-0.0779	Non-Antigen	-0.46908	Non-Allergen	Non-Toxin
QLLKQLFKK	0.8099	-0.4017	Non-Antigen	-0.37104	Non-Allergen	Non-Toxin
RAIDSQNMTK	0.752015	0.9180	Antigen	-0.33999	Allergen	Non-Toxin

Table (4) shows Cytotoxic T Lymphocyte (CTL) epitopes prediction for TcpF protein with their scores and antigenicity, immunogenicity, allergenicity and toxicity predictions for each epitope:

Protein2_TcpF						
Epitope	Score	Prediction for the Protective Antigen	Probable	Immunogenicity score	Allergenicity	Toxicity
KVYEGTLSRLK	0.989991	0.6767	Antigen	0.04102	Non-Allergen	Non-Toxin
NLWDVQFKV	0.988106	1.7075	Antigen	-0.00182	Non-Allergen	Non-Toxin
ATDSRGSEHLRY	0.981091	2.0337	Antigen	-0.08814	Non-Allergen	Non-Toxin

IPNEIYPHI	0.974817	0.8650	Antigen	0.23431	Allergen	Non-Toxin
VPDNTPVKL	0.950358	0.8256	Antigen	-0.06311	Non-Allergen	Non-Toxin

Table (5) shows Cytotoxic T Lymphocyte (CTL) epitopes prediction for OmpU protein with their scores and antigenicity, immunogenicity, allergenicity and toxicity predictions for each epitope:

Protein3_OmpU						
Epitope	Score	Prediction for the Protective Antigen	Probable	Immunogenicity score	Allergenicity	Toxicity
YSDNGEDGY	0.988818	1.2451	Antigen	0.16646	Non-Allergen	Non-Toxin
ITDFTDIMSY	0.985306	-0.7216	Non-Antigen	0.06482	Non-Allergen	Non-Toxin
VTETNAAKY	0.979829	1.0848	Antigen	0.00911	Non-Allergen	Non-Toxin
FAIDATYYF	0.975141	0.4178	Non-Antigen	0.13488	Allergen	Non-Toxin
YADQDDQNEY	0.948791	0.8233	Antigen	-0.12216	Non-Allergen	Non-Toxin

Table (6) shows Cytotoxic T Lymphocyte (CTL) epitopes prediction for OmpW protein with their scores and antigenicity, immunogenicity, allergenicity and toxicity predictions for each epitope:

Protein4_OmpW						
Epitope	Score	Prediction for the Protective Antigen	Probable	Immunogenicity score	Allergenicity	Toxicity
YANIETTATY	0.960733	0.6398	Antigen	0.35936	Allergen	Non-Toxin
YYFGEANSTF	0.960026	0.2983	Non-Antigen	0.08467	Allergen	Non-Toxin
LPPTFMVQY	0.925373	1.2261	Antigen	-0.04868	Allergen	Non-Toxin

YMLNDSWFL	0.905352	0.5992	Antigen	0.1111	Non-Allergen	Non-Toxin
ATYKAGADAK	0.851074	1.7833	Antigen	-0.06869	Allergen	Non-Toxin

Table (7) shows Cytotoxic T Lymphocyte (CTL) epitopes prediction for OmpT protein with their scores and antigenicity, immunogenicity, allergenicity and toxicity predictions for each epitope:

Protein5_OmpT						
Epitope	Score	Prediction for the Protective Antigen	Probable	Immunogenicity score	Allergenicity	Toxicity
VYLGVEYKF	0.993023	0.9671	Antigen	0.03583	Allergen	Non-Toxin
TTDIKADVTNSY	0.988568	0.8962	Antigen	-0.03456	Allergen	Non-Toxin
LSDALHDSQVKY	0.965197	0.4882	Non-Antigen	-0.27243	Non-Allergen	Non-Toxin
VYGADYSYF	0.961338	0.2495	Non-Antigen	-0.07329	Allergen	Non-Toxin
TSDDVYGADY	0.959666	1.4331	Antigen	0.144	Allergen	Non-Toxin

Table (8) shows Cytotoxic T Lymphocyte (CTL) epitopes prediction for AcfA protein with their scores and antigenicity, immunogenicity, allergenicity and toxicity predictions for each epitope:

Protein6_AcfA						
Epitope	Score	Prediction for the Protective Antigen	Probable	Immunogenicity score	Allergenicity	Toxicity
YSYPIHQQL	0.983821	0.5485	Antigen	-0.01775	Allergen	Non-Toxin
ALLETGLKK	0.924198	-0.0014	Non-Antigen	0.03149	Allergen	Non-Toxin

SLNNQQYRK	0.853936	1.1875	Antigen	-0.20333	Non-Allergen	Non-Toxin
LALISVYSY	0.843982	0.5355	Antigen	-0.0917	Non-Allergen	Non-Toxin
QFDKYNDVL	0.79284	-0.0592	Non-Antigen	-0.17665	Non-Allergen	Non-Toxin

Table (9) shows Cytotoxic T Lymphocyte (CTL) epitopes prediction for NlpD protein with their scores and antigenicity, immunogenicity, allergenicity and toxicity predictions for each epitope:

Protein7_NlpD						
Epitope	Score	Prediction for the Protective Antigen	Probable	Immunogenicity score	Allergenicity	Toxicity
LPNYTPPAY	0.996271	0.5967	Antigen	0.03504	Non-Allergen	Non-Toxin
ATADGTVVY	0.95354	1.1338	Antigen	0.16352	Non-Allergen	Non-Toxin
LTDKDVNDLISY	0.926054	0.2331	Non-Antigen	-0.12748	Allergen	Non-Toxin
STNSQNLT	0.89137	1.5023	Antigen	-0.27414	Allergen	Non-Toxin
TVDQTKTKEY	0.888689	1.2143	Antigen	-0.36852	Non-Allergen	Non-Toxin

Table (10) shows Cytotoxic T Lymphocyte (CTL) epitopes prediction for TolC protein with their scores and antigenicity, immunogenicity, allergenicity and toxicity predictions for each epitope:

Protein8_TolC						
Epitope	Score	Prediction for the Protective Antigen	Probable	Immunogenicity score	Allergenicity	Toxicity
AENSLTNSY	0.993123	0.7812	Antigen	-0.24495	Non-Allergen	Non-Toxin
QELYQRSSW	0.980323	0.1580	Non-Antigen	-0.30768	Allergen	Non-Toxin

ARYDYILSV	0.973194	0.8723	Antigen	0.03678	Allergen	Non-Toxin
VLDATRRLY	0.88554	-1.0413	Non-Antigen	0.17029	Non-Allergen	Non-Toxin
KLLPLFVSA	0.856946	0.9597	Antigen	0.02282	Allergen	Non-Toxin

Table (11) shows Cytotoxic T Lymphocyte (CTL) epitopes prediction for FlgO protein with their scores and antigenicity, immunogenicity, allergenicity and toxicity predictions for each epitope:

Protein9_FlgO						
Epitope	Score	Prediction for the Protective Antigen	Probable	Immunogenicity score	Allergenicity	Toxicity
RSDPTITQPY	0.972106	1.2277	Antigen	0.09574	Non-Allergen	Non-Toxin
LEMKRMKKW	0.9517	-0.1194	Non-Antigen	-0.6969	Allergen	Non-Toxin
LSSEQEVQY	0.873309	0.7758	Antigen	-0.00434	Non-Allergen	Non-Toxin
SEQEVQYVL	0.865841	0.5546	Antigen	0.01531	Non-Allergen	Non-Toxin
EPYNGSQFM	0.842787	-0.1868	Non-Antigen	-0.1598	Non-Allergen	Non-Toxin

Table (12) shows Cytotoxic T Lymphocyte (CTL) epitopes prediction for FlgP protein with their scores and antigenicity, immunogenicity, allergenicity and toxicity predictions for each epitope:

Protein10_FlgP						
Epitope	Score	Prediction for the Protective Antigen	Probable	Immunogenicity score	Allergenicity	Toxicity
KMRDYGEVQQV	0.83234	0.2038	Non-Antigen	0.03944	Allergen	Non-Toxin
AEQVYGMRI	0.616538	1.5490	Antigen	-0.08572	Allergen	Non-Toxin

SKIDAYREL	0.609704	-0.2674	Non-Antigen	0.20232	Allergen	Non-Toxin
MRISGRAEL	0.568598	0.7293	Antigen	0.04997	Allergen	Non-Toxin
VPEKRQMTL	0.554652	0.0482	Non-Antigen	-0.36866	Allergen	Non-Toxin

Table (13) shows Cytotoxic T Lymphocyte (CTL) epitopes prediction for FlaA protein with their scores and antigenicity, immunogenicity, allergenicity and toxicity predictions for each epitope:

Protein11_FlaA						
Epitope	Score	Prediction for the Protective Antigen	Probable	Immunogenicity score	Allergenicity	Toxicity
AEQPKTKEW	0.99844	0.5502	Antigen	-0.34572	Allergen	Non-Toxin
ALNEESVAL	0.943569	0.4573	Non-Antigen	0.09812	Non-Allergen	Non-Toxin
TYINGQTDLF	0.919833	-0.1409	Non-Antigen	0.00567	Allergen	Non-Toxin
VSAMTAQRY	0.861948	0.4867	Non-Antigen	-0.15689	Allergen	Non-Toxin
MTAQRYLTK	0.83927	0.1502	Non-Antigen	-0.04362	Allergen	Non-Toxin

The accepted epitopes were the epitopes which met the required conditions and whose binding score was higher than 0.9. The accepted epitopes are shown in the table (14) below:

Table (14) shows the accepted epitopes which met the required conditions and whose binding score was higher than 0.9:

Protein	TcpF	OmpU		OmpW	NlpD		FlgO
Epitopes	KVYEGTLSRLK	YSDNGEDGY	VTETNAAKY	YMLNDSWFL	LPNYTPPAY	ATADGTVVY	RSDPTITQPY
Score	0.989991	0.988818	0.979829	0.905352	0.996271	0.953543	0.972106

4-2-2- Helper T Lymphocyte (HTL) epitope prediction:

The most common alleles in the Syrian population for the DRB1 locus were DRB1*11 (26.4%), DRB1*04 (14%), and DRB1*07 (12%). However, the most frequent alleles for the DQB1 locus were DQB1*03 (40.9%) and DQB1*05 (25.1%) [35]. HTL epitopes were predicted by using the Immune Epitope Database Analysis Resource MHC class II prediction using NN-align [45]. The best 5 results for each protein (according to the fact that low ic50 = good binders considered that the repeated core sequences were considered one epitope) were subjected to antigenicity, allergenicity and toxicity predictions as well as the IFN-γ prediction by IFN-γ epitope server. The results are shown in the tables (15-25):

Table (15) shows Helper T Lymphocyte (HTL) epitopes prediction for TcpA protein with their IC50 and antigenicity, allergenicity, INF-γ inducibility and toxicity predictions for each epitope:

Protein1_TcpA						
Epitope	IC50	Prediction for the Protective Antigen	Probable	Allergenicity	IFN-γ prediction	Toxicity
LGIMGVVSAGVVTLA	3.40	0.1882	Non-Antigen	Non-Allergen	-0.097449458	Non-Toxin
EKLCTGTAPFTVAFG	4.80	0.5332	Antigen	Non-Allergen	0.080055581	Non-Toxin
DFETSVADAATGAGV	5.40	1.1857	Antigen	Allergen	-0.16163612	Non-Toxin
SVADAATGAGVIKSI	5.40	0.7101	Antigen	Non-Allergen	0.20965309	Non-Toxin
KAFAITVGGLTQAQC	6.40	0.5718	Antigen	Allergen	0.30998862	Non-Toxin

Table (16) shows Helper T Lymphocyte (HTL) epitopes prediction for TcpF protein with their IC50 and antigenicity, allergenicity, INF-γ inducibility and toxicity predictions for each epitope:

Protein2_TcpF						
Epitope	IC50	Prediction for the Protective Antigen	Probable	Allergenicity	IFN-γ prediction	Toxicity
NYSSTSTVYATSNEA	2.70	0.5905	Antigen	Allergen	0.016457318	Non-Toxin

RGLMGTTSVVNAIPN	4.70	-0.0288	Non-Antigen	Allergen	0.076341581	Non-Toxin
FAFNDNYSSTSTVYA	13	0.7836	Antigen	Non-Allergen	-0.23789942	Non-Toxin
MRYKKTLMLSIMITS	19.20	0.3968	Non-Antigen	Non Allergen	-0.25842171	Non-Toxin
DSLQKLYIDFYLAQT	28.60	0.7559	Antigen	Allergen	-0.55263763	Non-Toxin

Table (17) shows Helper T Lymphocyte (HTL) epitopes prediction for OmpU protein with their IC50 and antigenicity, allergenicity, INF-γ inducibility and toxicity predictions for each epitope:

Protein3_OmpU						
Epitope	IC50	Prediction for the Protective Antigen	Probable	Allergenicity	IFN-γ prediction	Toxicity
LIALAVSAAAVATGA	1.70	0.7088	Antigen	Non-Allergen	0.70255433	Non-Toxin
SAAAVATGAYADGIN	3.60	0.7538	Antigen	Non-Allergen	0.13836599	Non-Toxin
MNKTLIALAVSAAAV	5.00	0.1796	Non-Antigen	Non-Allergen	0.56890799	Non-Toxin
KPNFRSYISYQFNLL	5.40	1.0314	Antigen	Non-Allergen	0.23839563	Non-Toxin
QDDQNEYMLAASYRM	7.60	0.4908	Non-Antigen	Allergen	0.12138828	Non-Toxin

Table (18) shows Helper T Lymphocyte (HTL) epitopes prediction for OmpW protein with their IC50 and antigenicity, allergenicity, INF-γ inducibility and toxicity predictions for each epitope:

Protein4_OmpW						
Epitope	IC50	Prediction for the Protective Antigen	Probable	Allergenicity	IFN-γ prediction	Toxicity
GLAVLAALSSAPVFA	5.50	0.2048	Non-Antigen	Non-Allergen	0.27465023	Non-Toxin

AVLAALSSAPVFAHQ	6	0.1299	Non-Antigen	Non-Allergen	-0.069391535	Non-Toxin
ANIETTATYKAGADA	7.80	1.3595	Antigen	Non-Allergen	-0.86702836	Non-Toxin
GDFIVRAGIASVVPN	9.10	0.1638	Non-Antigen	Allergen	0.4977305	Non-Toxin
FTDNISFEVLAATPF	10.30	0.6785	Antigen	Non-Allergen	0.07635487	Non-Toxin

Table (19) shows Helper T Lymphocyte (HTL) epitopes prediction for OmpT protein with their IC50 and antigenicity, allergenicity, INF-γ inducibility and toxicity predictions for each epitope:

Protein5_OmpT						
Epitope	IC50	Prediction for the Protective Antigen	Probable	Allergenicity	IFN-γ prediction	Toxicity
AVLAAAGSVNAAEIL	3.40	0.3889	Non-Antigen	Non-Allergen	0.21510361	Non-Toxin
AAAGSVNAAEILKSD	9.30	0.6546	Antigen	Allergen	0.16497504	Non-Toxin
MKKTLLALAVLAAAG	9.60	-0.0238	Non-Antigen	Non-Allergen	0.47360123	Non-Toxin
DPTIGSGSSRAGVDA	9.90	2.0529	Antigen	Non-Allergen	-0.056728539	Non-Toxin
KRQWINMKKTLLALA	10.80	-0.6826	Non-Antigen	Allergen	0.10214279	Non-Toxin

Table (20) shows Helper T Lymphocyte (HTL) epitopes prediction for AcfA protein with their IC50 and antigenicity, allergenicity, INF-γ inducibility and toxicity predictions for each epitope:

Protein6_AcfA

Epitope	IC50	Prediction for the Protective Antigen	Probable	Allergenicity	IFN-γ prediction	Toxicity
IFLFTTLSANAAPYI	2.90	0.4413	Non-Antigen	Non-Allergen	-0.53479045	Non-Toxin
EAKQLALISVYSYPI	7.00	0.5869	Antigen	Non-Allergen	0.34207118	Non-Toxin
TTLSANAAPYIGLEL	9.00	0.8290	Antigen	Non-Allergen	-0.040333906	Non-Toxin
LSAIFLFTTLSANAA	10.00	0.5158	Antigen	Non-Allergen	-0.1913288	Non-Toxin
YSKFIGIESLNNQQY	18.20	-0.2143	Non-Antigen	Non-Allergen	-0.9017479	Non-Toxin

Table (21) shows Helper T Lymphocyte (HTL) epitopes prediction for NlpD protein with their IC50 and antigenicity, allergenicity, INF-γ inducibility and toxicity predictions for each epitope:

Protein7_NlpD						
Epitope	IC50	Prediction for the Protective Antigen	Probable	Allergenicity	IFN-γ prediction	Toxicity
VAVASSTSASVAKAA	1.90	0.8343	Antigen	Non-Allergen	0.18952027	Non-Toxin
SVAKAATTATVAQTV	1.90	0.5211	Antigen	Non-Allergen	0.69236818	Non-Toxin
AYGGTGGAATVAVAS	2.40	1.6544	Antigen	Non-Allergen	0.35326421	Non-Toxin
GQAVVATADGTVVYS	10.50	0.5643	Antigen	Non-Allergen	0.26506717	Non-Toxin
KIATMGSSGTNSVRL	10.70	1.2795	Antigen	Allergen	-0.32496754	Non-Toxin

Table (22) shows Helper T Lymphocyte (HTL) epitopes prediction for TolC protein with their IC50 and antigenicity, allergenicity, INF-γ inducibility and toxicity predictions for each epitope:

Protein8_TolC

Epitope	IC50	Prediction for the Protective Antigen	Probable	Allergenicity	IFN-γ prediction	Toxicity
VSAALGTLSSAVWAE	3.90	0.2055	Non-Antigen	Non-Allergen	0.53212175	Non-Toxin
AALGTLSSAVWAENL	4.00	0.1523	Non-Antigen	Allergen	0.18860259	Non-Toxin
AFEAVTSSRSALLPQ	4.30	0.2962	Non-Antigen	Allergen	-0.33061579	Non-Toxin
SALLPQINLTAGYNI	6.00	1.0594	Antigen	Non-Allergen	-0.6159339	Non-Toxin
LPLFVSAALGTLSSA	6.70	0.5368	Antigen	Non-Allergen	-0.084417645	Non-Toxin

Table (23) shows Helper T Lymphocyte (HTL) epitopes prediction for FlgO protein with their IC50 and antigenicity, allergenicity, INF-γ inducibility and toxicity predictions for each epitope:

Protein9_FlgO						
Epitope	IC50	Prediction for the Protective Antigen	Probable	Allergenicity	IFN-γ prediction	Toxicity
FKVVDFKTTGSIQVT	2.90	1.3383	Antigen	Allergen	-0.30028632	Non-Toxin
NGSQFMLMESPRHTL	3.40	0.1200	Non-Antigen	Non-Allergen	0.020519332	Non-Toxin
PVVLLTSCAYAPIYN	5.80	0.2408	Non-Antigen	Allergen	-0.10218892	Non-Toxin
IVYLEMKRMKKWLSL	6.10	-0.2346	Non-Antigen	Allergen	0.050398694	Non-Toxin
VYLEMKRMKKWLSLV	6.50	-0.2629	Non-Antigen	Allergen	0.049014365	Non-Toxin

Table (24) shows Helper T Lymphocyte (HTL) epitopes prediction for FlgP protein with their IC50 and antigenicity, allergenicity, INF-γ inducibility and toxicity predictions for each epitope:

Protein10_FlgP

Epitope	IC50	Prediction for the Protective Antigen	Probable	Allergenicity	IFN-γ prediction	Toxicity
RLGTELTAGAVDGVI	4.80	0.9082	Antigen	Allergen	0.13321535	Non-Toxin
DQRLGTELTAGAVDG	13.20	1.3561	Antigen	Allergen	-0.097824909	Non-Toxin
DDWLTAVGYANISEQ	29.30	-0.0712	Non-Antigen	Allergen	0.20183843	Non-Toxin
DWLTAVGYANISEQR	30.00	-0.0075	Non-Antigen	Allergen	0.37360216	Non-Toxin
LLLVAALMMTGCQPL	31.70	0.6160	Antigen	Non-Allergen	-0.28884172	Non-Toxin

Table (25) shows Helper T Lymphocyte (HTL) epitopes prediction for FlaA protein with their IC50 and antigenicity, allergenicity, INF-γ inducibility and toxicity predictions for each epitope:

Protein11_FlaA						
Epitope	IC50	Prediction for the Protective Antigen	Probable	Allergenicity	IFN-γ prediction	Toxicity
FQIGSSSGEAIIMGL	8.20	0.9201	Antigen	Non-Allergen	-0.10473343	Non-Toxin
ILQQAGTSILAQAKQ	8.40	-0.2717	Non-Antigen	Non-Allergen	-0.0016193658	Non-Toxin
LQQAGTSILAQAKQL	8.50	-0.1849	Non-Antigen	Non-Allergen	-0.21058299	Non-Toxin
INVNTNVSAMTAQRY	10.40	0.9958	Antigen	Allergen	0.034895713	Non-Toxin
ASFQIGSSSGEAIIM	10.60	1.0116	Antigen	Non-Allergen	-0.40956404	Non-Toxin

The accepted epitopes were entered into IL4pred which is an In-Silico platform for designing and discovering of Interleukin-4 inducing peptides. IL4pred allows users to predict whether their peptide has the ability to induce IL4 or not [48] IL4 inducing prediction results are shown in the Table (26).

Table (26) shows IL4 inducibility prediction for the best HTL epitopes:

Protein	TcpA		OmpU			OmpW	AcfA	NlpD			
IL4 inducing prediction	Non-inducer	Non-inducer	Non-inducer	Non-inducer	Inducer	Inducer	Inducer	Inducer	Non-inducer	Non-inducer	Inducer
Epitopes	EKLCTGTAPFTVAFG	SVADAATGAGVIKSI	LIALAVSAAAVATGA	SAAAVATGAYADGIN	KPNFRSYISYQFNLL	FTDNISFEVLAATPF	EAKQLALISVYSYPI	VAVASSTSASVAKAA	SVAKAATTATVAQTV	AYGGTGGAATVAVAS	GQAVVATADGTVVYS
IC50	4.80	5.40	1.70	3.60	5.40	10.30	7.00	1.90	1.90	2.40	10.50

The final accepted epitopes are shown in the Table (27)

Table (27) shows the final accepted HTL epitopes:

Protein	OmpU	OmpW	AcfA	NlpD	
IL4 inducing prediction	IL4-inducer	IL4-inducer	IL4-inducer	IL4-inducer	IL4-inducer
Epitopes	KPNFRSYISYQFNLL	FTDNISFEVLAATPF	EAKQLALISVYSYPI	VAVASSTSASVAKAA	GQAVVATADGTVVYS
IC50	5.40	10.30	7.00	1.90	10.50

4-2-3- Linear B Lymphocyte (LBL) epitope prediction:

ABCpred server [49] [50] was used in the LBL epitopes predictions. The amino acid sequence of vaccine candidates was submitted for 16mer B cell linear epitope prediction. Threshold was set at 0.51. The predicted B cell epitopes are ranked according to their score obtained by trained recurrent neural networks. Higher score of the peptide means the higher probability to be an epitope. Antigenicity, Immunogenicity, allergenicity and toxicity were predicted for the LBL epitopes. The results are shown in the Tables (28-38):

Table (28) shows Linear B Lymphocyte (LBL) epitopes prediction for TcpA protein with their scores and antigenicity, allergenicity and toxicity predictions for each epitope:

Protein1_TcpA

Epitope	Score	Prediction for the Protective Antigen	Probable	Allergenicity	Toxicity
RSLGNYPATANANAAT	0.93	0.5579	Antigen	Non-Allergen	Non-Toxin
AGVIKSIAPGSANLNL	0.92	0.3801	Non-Antigen	Allergen	Non-Toxin
SVQIAMTQTYRSLGNY	0.91	0.1219	Non-Antigen	Allergen	Non-Toxin
PFTGTAMGIFSFPRNS	0.91	-0.3325	Non-Antigen	Non-Allergen	Non-Toxin
GLVSLGKVSADEAKNP	0.83	-0.1574	Non-Antigen	Allergen	Non-Toxin

Table (29) shows Linear B Lymphocyte (LBL) epitopes prediction for TcpF protein with their scores and antigenicity, allergenicity and toxicity predictions for each epitope:

Protein2_TcpF					
Epitope	Score	Prediction for the Protective Antigen	Probable	Allergenicity	Toxicity
CIKIGMSRDYLENCVK	0.92	0.2627	Non-Antigen	Allergen	Non-Toxin
TGVIYDPVYEETVKPY	0.91	0.0312	Non-Antigen	Non-Allergen	Non-Toxin
RGLMGTTSVVNAIPNE	0.91	0.0052	Non-Antigen	Allergen	Non-Toxin
VSTNDMHNGYKWSNTM	0.89	0.4512	Non-Antigen	Non-Allergen	Toxin
DWEIPTRDQIETLVNY	0.88	-0.0516	Non-Antigen	Non-Allergen	Non-Toxin

Table (30) shows Linear B Lymphocyte (LBL) epitopes prediction for OmpU protein with their scores and antigenicity, allergenicity and toxicity predictions for each epitope:

Protein3_OmpU

Epitope	Score	Prediction for the Protective Antigen	Probable	Allergenicity	Toxicity
AGIGGTYGEVTYGKND	0.93	1.1639	Antigen	Allergic	Non-Toxin
GFYEGEFTTNDQGKNA	0.88	0.4979	Non-Antigen	Non-Allergen	Non-Toxin
AASYRMENLYFAGLFT	0.88	0.3296	Non-Antigen	Allergen	Non-Toxin
GQAAFTATYNNAETAK	0.87	0.8086	Antigen	Allergic	Non-Toxin
TGFNVGAGYADQDDQN	0.84	0.8005	Antigen	Allergic	Non-Toxin

Table (31) shows Linear B Lymphocyte (LBL) epitopes prediction for OmpW protein with their scores and antigenicity, allergenicity and toxicity predictions for each epitope:

Protein4_OmpW					
Epitope	Score	Prediction for the Protective Antigen	Probable	Allergenicity	Toxicity
YANIETTATYKAGADA	0.95	1.1540	Antigen	Allergic	Non-Toxin
NGTGTNAGLSDLKLDD	0.92	2.0919	Antigen	Non-Allergen	Non-Toxin
DVEINPWVFMIAGGYK	0.90	1.3282	Antigen	Non-Allergen	Non-Toxin
SWGLAANVGFDYMLND	0.89	1.1164	Antigen	Allergic	Non-Toxin
GETKHLPPTFMVQYYF	0.89	1.2222	Antigen	Allergic	Non-Toxin

Table (32) shows Linear B Lymphocyte (LBL) epitopes prediction for OmpT protein with their scores and antigenicity, allergenicity and toxicity predictions for each epitope:

Protein5_OmpT

Epitope	Score	Prediction for the Protective Antigen	Probable	Allergenicity	Toxicity
SGSSRAGVD ANYTVND	0.93	1.5795	Antigen	Allergic	Non-Toxin
NSHIKKRQW INMKKTL	0.91	0.0657	Non-Antigen	Non-Allergen	Non-Toxin
EYTIGDALIG VTYYNA	0.89	0.5992	Antigen	Non-Allergen	Non-Toxin
YVMQEANT GADEDGTL	0.88	1.1547	Antigen	Non-Allergen	Non-Toxin
TKLYAGYEY VMQEANT	0.88	0.3541	Non-Antigen	Allergen	Non-Toxin

Table (33) shows Linear B Lymphocyte (LBL) epitopes prediction for AcfA protein with their scores and antigenicity, allergenicity and toxicity predictions for each epitope:

Protein6_AcfA					
Epitope	Score	Prediction for the Protective Antigen	Probable	Allergenicity	Toxicity
SVGTELRYQ FDKYNDV	0.85	1.0308	Antigen	Allergic	Non-Toxin
THTQYEAYS GKYEELE	0.885	0.8585	Antigen	Non-Allergen	Non-Toxin
LETGLKKNR FGALFSL	0.84	0.6055	Antigen	Non-Allergen	Non-Toxin
LSAIFLFTTL SANAAP	0.82	0.4355	Non-Antigen	Non-Allergen	Non-Toxin
LGIGTANHS FETNYQS	0.78	0.7288	Antigen	Non-Allergen	Non-Toxin

Table (34) shows Linear B Lymphocyte (LBL) epitopes prediction for NlpD protein with their scores and antigenicity, allergenicity and toxicity predictions for each epitope:

Protein7_NlpD					
Epitope	Score	Prediction for the Protective Antigen	Probable	Allergenicity	Toxicity
GGTGGAAT VAVASSTS	0.94	1.8360	Antigen	Non-Allergen	Non-Toxin
EVKKGDTLY FIAYLTD	0.93	0.3473	Non-Antigen	Non-Allergen	Non-Toxin
DEKIAKWL WPTKGRVI	0.93	0.0875	Non-Antigen	Allergen	Non-Toxin
GSYRGSYYE VKKGDTL	0.92	0.7434	Antigen	Allergic	Non-Toxin
EIRYQGKSV NPKRYLP	0.88	0.7665	Antigen	Allergic	Non-Toxin

Table (35) shows Linear B Lymphocyte (LBL) epitopes prediction for TolC protein with their scores and antigenicity, allergenicity and toxicity predictions for each epitope:

Protein8_TolC					
Epitope	Score	Prediction for the Protective Antigen	Probable	Allergenicity	Toxicity
DVLDATRRL YDANKNL	0.93	-0.4998	Non-Antigen	Non-Allergen	Non-Toxin
VGLSAITDV HDAQAQF	0.93	0.5451	Antigen	Allergic	Non-Toxin
NTSGEEYND FKIGVNL	0.92	1.6409	Antigen	Allergic	Non-Toxin
IGTLSEQDV MDVNAGL	0.90	0.3640	Non-Antigen	Non-Allergen	Non-Toxin
RQLEQTKQR FEVGLSA	0.89	0.9526	Antigen	Non-Allergen	Non-Toxin

Table (36) shows Linear B Lymphocyte (LBL) epitopes prediction for FlgO protein with their scores and antigenicity, allergenicity and toxicity predictions for each epitope:

Protein9_FlgO					
Epitope	Score	Prediction for the Protective Antigen	Probable	Allergenicity	Toxicity
TSCAYAPIY NGKEPYN	0.90	0.1239	Non-Antigen	Allergen	Non-Toxin
TEDLMLSNT SITARTP	0.87	0.6801	Antigen	Non-Allergen	Non-Toxin
GKEPYNGSQ FMLMESP	0.86	0.2632	Non-Antigen	Non-Allergen	Non-Toxin
GVLIRSDPTI TQPYTV	0.85	0.6689	Antigen	Non-Allergen	Non-Toxin
TQQGDFAFS RDWKNLS	0.84	0.0103	Non-Antigen	Non-Allergen	Non-Toxin

Table (37) shows Linear B Lymphocyte (LBL) epitopes prediction for FlgP protein with their scores and antigenicity, allergenicity and toxicity predictions for each epitope:

Protein10_FlgP					
Epitope	Score	Prediction for the Protective Antigen	Probable	Allergenicity	Toxicity
AEVVRSYKV GDSYVTE	0.93	0.9579	Antigen	Allergic	Non-Toxin
SMRPDDWL TAVGYANI	0.91	0.1886	Non-Antigen	Allergen	Non-Toxin
TELRLDIRK MDKMRDY	0.90	0.7626	Antigen	Non-Allergen	Non-Toxin
MMTGCQPL QSMRPDDW	0.89	0.4752	Non-Antigen	Non-Allergen	Non-Toxin
EQVYGMRIS GRAELQD	0.86	1.0983	Antigen	Non-Allergen	Non-Toxin

Table (38) shows Linear B Lymphocyte (LBL) epitopes prediction for FlaA protein with their scores and antigenicity, allergenicity and toxicity predictions for each epitope:

Protein11_FlaA					
Epitope	Score	Prediction for the Protective Antigen	Probable	Allergenicity	Toxicity
GDDIEELATYINGQTD	0.91	0.3691	Non-Antigen	Non-Allergen	Non-Toxin
PKTKEWGVPPTARDLK	0.90	0.0366	Non-Antigen	Non-Allergen	Non-Toxin
GEAIIMGLTSVRADDF	0.90	0.3382	Non-Antigen	Allergen	Non-Toxin
TKATGELNTSMERLSS	0.88	0.7503	Antigen	Allergic	Non-Toxin
SFIAEQPKTKEWGVPP	0.88	0.1119	Non-Antigen	Non-Allergen	Non-Toxin

The epitopes which met the required conditions and whose binding score was higher than 0.9 were accepted. The accepted epitopes are shown in the table (39):

Table (39) shows the LBL accepted epitopes:

Protein	TcpA	OmpW		OmpT	AcfA	NlpD	TolC	FlgP
Epitopes	RSLGNYPATANANAAT	NGTGTNAGLSDLKLDD	DVEINPWVFMIAGGYK	EYTIGDALIGVTYYNA	THTQYEAYSGKYEELE	GGTGGAATVAVASSTS	RQLEQTKQRFEVGLSA	TELRLDIRKMDKMRDY
Sore	0.93	0.92	0.90	0.89	0.885	0.94	0.89	0.90

The final accepted epitopes are shown in the tables (40-41-42):

Table (40) shows the final accepted CTL epitopes:

Protein	TcpF	OmpU		OmpW	NlpD		FlgO
Epitopes	KVYEGTLSRLK	YSDNGEDGY	VTETNAAKY	YMLNDSWFL	LPNYTPPAY	ATADGTVVY	RSDPTITQPY

Score	0.989991	0.988818	0.979829	0.905352	0.996271	0.953543	0.972106
Allele	HLA-A*03	HLA-A*01	HLA-A*01	HLA-A*02	HLA-B*35	HLA-A*01	HLA-A*01

Table (41) shows the final accepted HTL epitopes:

Protein	OmpU	OmpW	AcfA	NlpD	
Epitopes	KPNFRSYISYQFNLL	FTDNISFEVLAATPF	EAKQLALISVYSYPI	VAVASSTSASVAKAA	GQAVVATADGTVVYS
IC50	5.40	10.30	7.00	1.90	10.50
Allele	DRB1*07	DRB1*07	DRB1*07	DQB1*03	DQB1*03

Table (42) shows the final accepted LBL epitopes:

Protein	TcpA	OmpW		OmpT	AcfA	NlpD	TolC	FlgP
Epitopes	RSLGNYPATANANAAT	NGTGTNAGLSDLKLDD	DVEINPWVFMIAGGYK	EYTIGDALIGVTYYNA	THTQYEAYSGKYEELE	GGTGGAATVAVASSTS	RQLEQTKQRFEVGLSA	TELRLDIRKMDKMRDY
Sore	0.93	0.92	0.90	0.89	0.885	0.94	0.89	0.90

4-3- Population Coverage Calculation:

A web-based tool from IEDB Analysis Resource was developed to predict population coverage of T-cell epitope-based diagnostics and vaccines based on MHC binding and/or T cell restriction data [51]. The aforementioned CTL and HTL accepted epitopes were entered into this tool separately. The population coverage calculation result for CTL epitopes is shown in the Table (43).

Table (43) shows population coverage calculation results for MHC class I epitopes:

population/area	Class I
	coverage
World	66.83%
Average	66.83

In regard to MHC I polymorphism, the Syrian population is genetically closer to neighboring human populations, (Jordanians, Lebanese, and Turks) and to Europeans in the north of the Mediterranean. The Syrians are genetically far from human populations from the Arabian Peninsula and North Africa, and very far from the Chinese and other African human populations. [34]. When the search was specialized to the aforementioned populations which are much closer to the Syrian population, the results was as shown in the table (44):

Table (44) shows population coverage calculation results for MHC class I epitopes when the search was specialized to populations close to the Syrian populations:

population/area	Class I
	coverage
Europe	81.4%
Jordan	46.3%
Turkey	44.8%
Turkey Caucasoid	44.8%

The population coverage calculation result for HTL epitopes is shown in the Table (45).

Table (45) shows population coverage calculation results for MHC class II epitopes:

population/area	Class II
	coverage
Europe	46.92%
Jordan	20.54%
Lebanon	62.97%
Turkey	54.16%
World	45.58%
Average	46.03
Standard deviation	14.17

In regard to MHC II polymorphism, the population of Syria is genetically closer to Lebanese, Jordanians, and Iranians compared with Austrians, Italians, and Koreans. The greater genetic distances from the Syrian population were seen in Chinese [35].

4-4- Vaccine construction:

The subunit B of Cholera toxin (CTB) is a good adjuvant [52]. The selection of appropriate linkers is an essential step in designing an immunogenic multi-epitope vaccine so that the domains can work independently avoiding interaction and interference between them [72]. Multiple combinations of the vaccine candidate (depending on studies in the medical literature) were proposed and discussed. And the best combination was approved as a final format for the vaccine. Another good combination was used for comparison.

4-4-1- The first combination (A):

CTL epitopes was joined by AAY linkers. AAY (Ala-Ala-Tyr) linker is the cleavage site for the proteasomes in mammalian cells. AAY linkers separate epitopes effectively, reduce the junctional immunogenicity and increase the immunogenicity of the multi-epitope vaccine. GPGPG linker was used in separating HTL epitopes. It is a valuable tool in breaking the junctional immunogenicity. KK linker was used to join the B-cell epitopes. The Lysine linker is the target for the Cathepsin B (lysosomal protease) and it plays a crucial role in reducing the junctional immunogenicity and increasing immunogenicity. CTB was added as an adjuvant using EAAAK linker. EAAAK is a rigid a-helix forming peptide linker [53].

The first combination was:

MIKLKFGVFFTVLLSSAYAHGTPQNITDLCAEYHNTQIYTLNDKIFSYTESLAGKREMAII
TFKNGAIFQVEVPGSQHIDSQKKAIERMKDTLRIAYLTEAKVEKLCVWNNKTPHAIAAIS
MANEAAAKKVYEGTLSRLKAAYYSDNGEDGYAAYVTETNAAKYAAYYMLNDSWFL
AAYLPNYTPPAYAAYATADGTVVYAAYRSDPTITQPYGPGPGKPNFRSYISYQFNLLGP
GPGFTDNISFEVLAATPFGPGPGEAKQLALISVYSYPIGPGPGVAVASSTSASVAKAAGP
GPGGQAVVATADGTVVYSKKRSLGNYPATANANAATKKNGTGTNAGLSDLKLDDKK
DVEINPWVFMIAGGYKKKEYTIGDALIGVTYYNAKKTHTQYEAYSGKYEELEKKGGTG
GAATVAVASSTSKKRQLEQTKQRFEVGLSAKKTELRLDIRKMDKMRDY

4-4-2-The second combination (B):

The epitopes were linked by a flexible GDGDG linker. At the end of this sequence, CTB was added as an adjuvant using EAAAK rigid linker[73].

The second combination was:

KVYEGTLSRLKGDGDGYSDNGEDGYGDGDGVTETNAAKYGDGDGYMLNDSWFLGDG
DGLPNYTPPAYGDGDGATADGTVVYGDGDGRSDPTITQPYGDGDGKPNFRSYISYQFN
LLGDGDGFTDNISFEVLAATPFGDGDGEAKQLALISVYSYPIGDGDGVAVASSTSASVA
KAAGDGDGGQAVVATADGTVVYSGDGDGRSLGNYPATANANAATGDGDGNGTGTN
AGLSDLKLDDGDGDGDVEINPWVFMIAGGYKGDGDGEYTIGDALIGVTYYNAGDGDG
THTQYEAYSGKYEELEGDGDGGGTGGAATVAVASSTSGDGDGRQLEQTKQRFEVGLS
AGDGDGTELRLDIRKMDKMRDYEAAAKMIKLKFGVFFTVLLSSAYAHGTPQNITDLCA
EYHNTQIYTLNDKIFSYTESLAGKREMAIITFKNGAIFQVEVPGSQHIDSQKKAIERMKDT
LRIAYLTEAKVEKLCVWNNKTPHAIAAISMAN

4-4-3-The third combination (C):

The epitopes were linked together by GPGPG linkers as well as CTB as an adjuvant was linked using EAAAK linker [14].

The third combination was:

MIKLKFGVFFTVLLSSAYAHGTPQNITDLCAEYHNTQIYTLNDKIFSYTESLAGKREMAII
TFKNGAIFQVEVPGSQHIDSQKKAIERMKDTLRIAYLTEAKVEKLCVWNNKTPHAIAAIS
MANEAAAKKVYEGTLSRLKGPGPGYSDNGEDGYGPGPGVTETNAAKYGPGPGYMLN
DSWFLGPGPGLPNYTPPAYGPGPGATADGTVVYGPGPGRSDPTITQPYGPGPGKPNFRS
YISYQFNLLGPGPGFTDNISFEVLAATPFGPGPGEAKQLALISVYSYPIGPGPGVAVASSTS
ASVAKAAGPGPGGQAVVATADGTVVYSGPGPGRSLGNYPATANANAATGPGPGNGTG
TNAGLSDLKLDDGPGPGDVEINPWVFMIAGGYKGPGPGEYTIGDALIGVTYYNAGPGPG
THTQYEAYSGKYEELEGPGPGGGTGGAATVAVASSTSGPGPGRQLEQTKQRFEVGLSA
GPGPGTELRLDIRKMDKMRDY

4-4-4- The fourth combination (D):

The epitopes were joined to each other with a three AAA linker. CTB protein was added as an adjuvant using EAAAK linker [74].

The fourth combination was:

MIKLKFGVFFTVLLSSAYAHGTPQNITDLCAEYHNTQIYTLNDKIFSYTESLAGKREMAII
TFKNGAIFQVEVPGSQHIDSQKKAIERMKDTLRIAYLTEAKVEKLCVWNNKTPHAIAAIS
MANEAAAKKVYEGTLSRLKAAAYSDNGEDGYAAAVTETNAAKYAAAYMLNDSWFL
AAALPNYTPPAYAAAATADGTVVYAAARSDPTITQPYAAAKPNFRSYISYQFNLLAAAF
TDNISFEVLAATPFAAAEAKQLALISVYSYPIAAAVAVASSTSASVAKAAAAGQAVVA
TADGTVVYSAAARSLGNYPATANANAATAAANGTGTNAGLSDLKLDDAAADVEINPW
VFMIAGGYKAAAEYTIGDALIGVTYYNAAATHTQYEAYSGKYEELEAAAGGTGGAA
TVAVASSTSAAARQLEQTKQRFEVGLSAAATELRLDIRKMDKMRDY

4-5- Physicochemical and other properties prediction:

The physicochemical properties of the vaccine candidate were assessed by ProtParam tool / Expasy [57]. Solubility was predicted by both Protein-Sol web tool [58] and SOLpro predictor [59]. The results are shown in the Table (46).

Table (46) shows physicochemical and solubility predictions for A, B, C and D proposed combinations:

Combination	A	B	C	D
Number of amino acids	457	493	493	455
Molecular weight	49395.94	51311.53	50628.60	47737.70
Theoretical pI	9.13	4.20	5.67	5.67
Total number of negatively charged residues (Asp + Glu)	41	79	41	41
Total number of positively charged residues (Arg + Lys):	52	36	36	36
Formula	C2225H3448N580O672S10	C2225H3388N600O778S10	C2263H3464N600O702S10	C2130H3312N562O664S10
Estimated half-life:	The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro). >20 hours (yeast, in vivo). >10 hours (Escherichia coli, in vivo).	The estimated half-life is: 1.3 hours (mammalian reticulocytes, in vitro). 3 min (yeast, in vivo). 3 min (Escherichia coli, in vivo).	The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro). >20 hours (yeast, in vivo). >10 hours (Escherichia coli, in vivo).	The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro). >20 hours (yeast, in vivo). >10 hours (Escherichia coli, in vivo).
Instability index:	27.63 (Stable)	21.17 (Stable)	20.66 (Stable)	26.56 (Stable)
Aliphatic index	72.74	64.99	64.99	82.95
Grand average of hydropathicity (GRAVY)	-0.333 (Hydrophilic)	-0.482 (Hydrophilic)	-0.335 (Hydrophilic)	0.046 (Hydrophobic)
Solubility (Protein-Sol)	0.420 INSOLUBLE	0.676 SOLUBLE	0.411 INSOLUBLE	0.383 INSOLUBLE

Solubility (SOLpro)	INSOLUBLE with probability 0.528632	INSOLUBLE with probability 0.687969	SOLUBLE with probability 0.855315	SOLUBLE with probability 0.768454

The antigenicity, allergenicity, homology with homo-sapiens proteins and toxicity of the proposed combinations were predicted. The results are shown in the Table (47):

Table (47) shows antigenicity, allergenicity, homology with homo-sapiens proteins and toxicity predictions of the the proposed combinations:

	A	B	C	D
Antigenicity (vaxijen)	0.8165 (Probable Antigen)	1.3112 (Probable Antigen)	0.9974 (Probable Antigen)	0.7578 (Probable Antigen)
Allergenicity (Allertop)	Probable Non-Allergen	Probable Non-Allergen	Probable Non-Allergen	Probable Non-Allergen
Homology with homo-sapiens proteins (blastp)	Non-Homologous	Non-Homologous	Non-Homologous	Non-Homologous
Toxicity (Toxinpred)	Non-toxin	Non-toxin	Non-toxin	Non-toxin

Combination (A): The molecular weight is predicted to be 49,395.94 kDa; with a theoretical pI 9.13 indicating its basic nature. The instability index of the vaccine is estimated to be 27.63, indicating that it is a stable protein. In addition, the aliphatic index and GRAVY score are 72.74 and -0.333, respectively, showing hydrophilic vaccine nature. Moreover, the immunological potency of the vaccine is determined by assessing the antigenicity score. The vaccine is antigenic, with a score of 0.8165. It is non-allergen.
The predicted solubility by Protein-Sol is 0.420 which indicates that the protein is insoluble. The protein is predicted to be insoluble with probability 0.528632 upon overexpression.

Combination (B): The molecular weight is predicted to be 51,311.53 kDa; with a theoretical pI 4.20 indicating its acidic nature. The instability index of the vaccine is estimated to be 20.66, indicating that it is a stable protein. In addition, the aliphatic index and GRAVY score are 64.99 and -0.482, respectively, showing hydrophilic vaccine nature. Moreover, the immunological potency of the vaccine is determined by assessing the antigenicity score. The vaccine is antigenic, with a score of 1.3112. It is non-allergen.
The predicted solubility by Protein-Sol is 0.676 which indicates that the protein is soluble. The protein is predicted to be insoluble with probability 0.687969 upon overexpression.

Combination (C): The molecular weight is predicted to be 50,628.60 kDa; with a theoretical pI 5.67 indicating its basic nature. The instability index of the vaccine is estimated to be 20.66, indicating that it is a stable protein. In addition, the aliphatic index and GRAVY score are 64.99 and -0.335, respectively, showing hydrophilic vaccine nature. Moreover, the immunological potency of the vaccine is determined by assessing the antigenicity score. The vaccine is antigenic, with a score of 0.9974. It is non-allergen.
The predicted solubility by Protein-Sol is 0.411 which indicates that the protein is insoluble. The protein is predicted to be soluble with probability 0.855315 upon overexpression.

Combination (D): The molecular weight is predicted to be 47,737.70 kDa; with a theoretical pI 5.67 indicating its basic nature. The instability index of the vaccine is estimated to be 26.56, indicating that it is a stable protein. In addition, the aliphatic index and GRAVY score are 82.95and 0.046, respectively, showing hydrophobic vaccine nature. Moreover, the immunological potency of the vaccine is determined by assessing the antigenicity score. The vaccine is antigenic, with a score of 0.7576. It is non-allergen.
The predicted solubility by Protein-Sol is 0.383 which indicates that the protein is insoluble. The protein is predicted to be soluble with probability 0.768454 upon overexpression.

The combination (B) was ignored because its estimated half-life is very short. Also, the combination (D) was ignored because of its positive GRAVY which indicated its hydrophobic nature. Combination (A) was adopted as a vaccine candidate. Combination (C) was used just for comparison, considering the basic nature for the combination (A) and the acidic nature for the combination (C). The insolubility of these combinations was discussed in more detail.

Adding solubility-enhancing peptide tags to less soluble proteins can improve protein solubility [61]. Solubility-enhancing peptide tags are short peptide tags comprising mostly one or two amino acids repeated a varying number of times. They are polar and bear a positive or negative overall charge [60]. They are suggested to assist correct protein folding and enhance solubility. The introduction of short peptide tag bearing similar charge as the protein of interest at a certain pH value in either of the protein's termini improves solubility due to inter- and intramolecular repulsive interactions and they less likely to disturb the active sites of target proteins, compromise its activity or interfere with the structure of the protein of interest [60] [61]. An extra step for the removal of the peptide tags after production and/or purification is not necessarily required [60]. Multiple proposed tags were tried and the solubility of the candidates (A) and (C) after adding the tags was predicted by Protein-sol and Solpro. The solubility predictions showed that adding ten residues of Lysine to the candidate (A) improved the solubility of this candidate and adding five residues of Glutamic acid to the candidate (C) improved the solubility of this candidate as it shown in the Table (48).

Table (48) shows solubility predictions through Protein-sol and SOLpro after adding tags to the candidates (A) and (C):

The added tag	A		C	
	Protein-sol	SOLpro	Protein-sol	SOLpro
(Arg)10	0.513 Soluble	INSOLUBLE with probability 0.507918	0.423 Insoluble	INSOLUBLE with probability 0.525364
(Lys)10	0.551 Soluble	SOLUBLE with probability 0.500000	0.439 Insoluble	SOLUBLE with probability 0.500000
(Glu)5	0.385 Insoluble	SOLUBLE with probability 0.517628	0.452 Soluble	SOLUBLE with probability 0.517628
(Asp)5	0.383 Insoluble	INSOLUBLE with probability 0.515148	0.452 Soluble	INSOLUBLE with probability 0.515148
(Arg-Gly-Gly)3-Gly	0.464 Soluble	INSOLUBLE with probability 0.525364	0.413 Insoluble	INSOLUBLE with probability 0.525364

New sequence of candidate (A):

MIKLKFGVFFTVLLSSAYAHGTPQNITDLCAEYHNTQIYTLNDKIFSYTESLAGKREMAII
TFKNGAIFQVEVPGSQHIDSQKKAIERMKDTLRIAYLTEAKVEKLCVWNNKTPHAIAAIS
MANEAAAKKVYEGTLSRLKAAYYSDNGEDGYAAYVTETNAAKYAAYYMLNDSWFL
AAYLPNYTPPAYAAYATADGTVVYAAYRSDPTITQPYGPGPGKPNFRSYISYQFNLLGP
GPGFTDNISFEVLAATPFGPGPGEAKQLALISVYSYPIGPGPGVAVASSTSASVAKAAGP
GPGGQAVVATADGTVVYSKKRSLGNYPATANANAATKKNGTGTNAGLSDLKLDDKK
DVEINPWVFMIAGGYKKKEYTIGDALIGVTYYNAKKTHTQYEAYSGKYEELEKKGGTG
GAATVAVASSTSKKRQLEQTKQRFEVGLSAKKTELRLDIRKMDKMRDYKKKKKKKKK
K

New sequence of candidate (C):

MIKLKFGVFFTVLLSSAYAHGTPQNITDLCAEYHNTQIYTLNDKIFSYTESLAGKREMAII
TFKNGAIFQVEVPGSQHIDSQKKAIERMKDTLRIAYLTEAKVEKLCVWNNKTPHAIAAIS
MANEAAAKKVYEGTLSRLKGPGPGYSDNGEDGYGPGPGVTETNAAKYGPGPGYMLN
DSWFLGPGPGLPNYTPPAYGPGPGATADGTVVYGPGPGRSDPTITQPYGPGPGKPNFRS
YISYQFNLLGPGPGFTDNISFEVLAATPFGPGPGEAKQLALISVYSYPIGPGPGVAVASSTS
ASVAKAAGPGPGGQAVVATADGTVVYSGPGPGRSLGNYPATANANAATGPGPGNGTG
TNAGLSDLKLDDGPGPGDVEINPWVFMIAGGYKGPGPGEYTIGDALIGVTYYNAGPGPG
THTQYEAYSGKYEELEGPGPGGGTGGAATVAVASSTSGPGPGRQLEQTKQRFEVGLSA
GPGPGTELRLDIRKMDKMRDYEEEEE

Physicochemical properties, antigenicity, allergenicity, homology with homo-sapiens proteins and toxicity predictions of the vaccine candidates after adding solubility-enhancing peptide tags was reassessed as shown in Table (49).

Table (49) shows physicochemical properties predictions after adding solubility-enhancing peptide tags:

	Candidate (A) after adding (Lysine)10	Candidate (C) after adding (Glutamic acid)5
Number of amino acids	467	498
Molecular weight	50677.68	51274.17
Theoretical pI	9.45	5.17
Total number of negatively charged residues (Asp + Glu)	41	46
Total number of positively charged residues (Arg + Lys):	62	36
Formula	C2285H3568N600O682S10	C2288H3499N605O717S10
Estimated half-life:	The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro). >20 hours (yeast, in vivo). >10 hours (Escherichia coli, in vivo)	The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro). >20 hours (yeast, in vivo). >10 hours (Escherichia coli, in vivo).
Instability index:	27.25 (Stable)	23.02 (Stable)
Aliphatic index	71.18	64.34
Grand average of hydropathicity (GRAVY)	-0.410	-0.367
Antigenicity (vaxijen)	0.8369 (Probable Antigen)	1.0004 (Probable Antigen)
Allergenicity (Allertop)	Probable Non-Allergen	Probable Non-Allergen

There were no changes in the physicochemical properties in both candidates after adding solubility-enhancing peptide tags.

4-6- Secondary structure:

The secondary structural features include a-helix, b-strand and random coils were evaluated using the SOPMA server [62] . The results are shown in the Table (50).

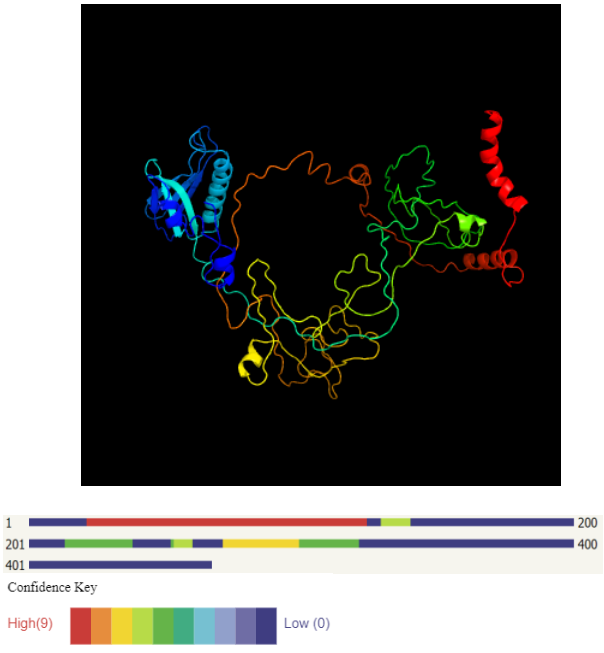
Table (50) shows secondary structure predictions for candidates (A) and (C) using SOPMA server:

	Candidate (A)	Candidate (C)
a-helix	34.90%	16.06%
Extended strands	24.84%	35.34%
Random coils	40.26%	48.59%

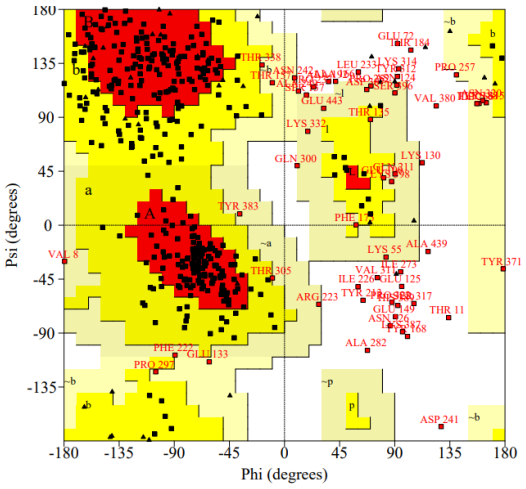
Vaccine candidate (A) was consisted of 34.90% a-helix, 24.84% extended strands and 40.26% random coils when vaccine candidate (C) was consisted of 16.06% a-helix, 35.34% extended strands and 48.59% random coils.

4-7- Tertiary structure, refinement and validation:

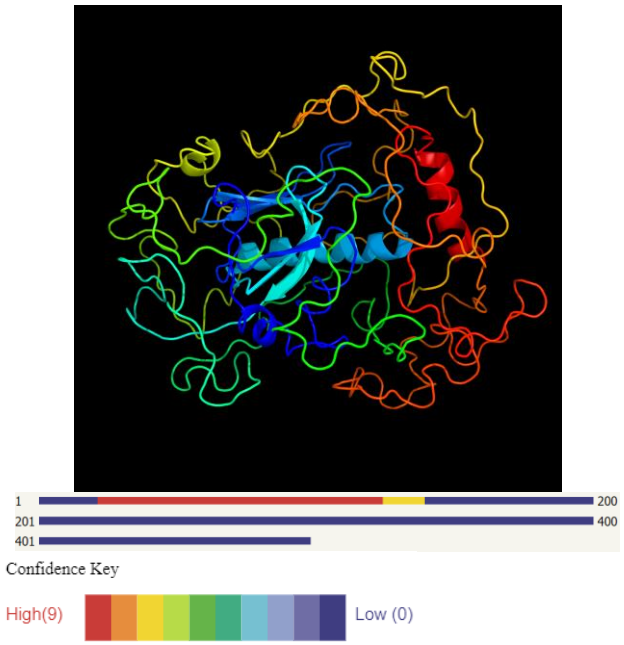
The Phyre2 server [63] was used in the homology modeling for both candidates using intensive modeling mode. Ramachandran plots for both candidates were created using Prochek [64]. Phyre2 models and Ramachandran plots are shown in Figure (1).

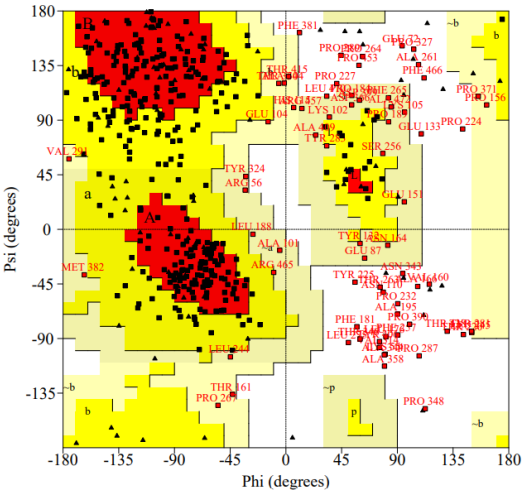


a



b





c **d**

Fig (1)

a: Vaccine candidate (A) protein modeled by the Phyre2 server. 22% of residues modeled at >90% confidence with the confidence of the model.
b: Ramachandran plot for the vaccine candidate (A) modeled protein shows: 60.7% of residues in most favored regions, 26.4% residues in additional allowed regions, 6.7% residues in generously allowed regions and 6.2% residues in disallowed regions, which indicates the low quality of the model.
c: Vaccine candidate (C) protein modeled by the Phyre2 server. 21% of residues modeled at >90% confidence with the confidence of the model. Warning: 76% of the sequence (C) was predicted to be disordered. Disordered regions cannot be meaningfully predicted.
d: Ramachandran plot for the vaccine candidate (C) modeled protein shows: 57.0% of residues in most favored regions, 27.8% residues in additional allowed regions, 8.3% residues in generously allowed regions and 6.9% residues in disallowed regions, which indicates the low quality of the model.

A good quality model would be expected to have over 90% in the most favored regions. As a result, Ramachandran plots indicated that there was a need for the refinement process by GalaxyRefine web server [75].

4-7-1- Candidate (A) refinement:

Galaxy-refine resulted in five models. They were evaluated depending on the Ramachandran plot statistical analysis by Prochek [64]. The obtained models shown in the Table (51):

Table (51) shows Ramachandran plot statistical analysis for the first refinement models for candidate (A) by Galaxy refine web server:

	Model 1	Model 2	Model 3	Model 4	Model 5
Residues in most favored regions	76.1%	76.4%	77.1%	76.1%	76.1%
Residues in additional allowed regions	17.9%	17.9%	16.9%	17.9%	17.9%
Residues in generously allowed regions	1.2 %	1.2 %	1.2 %	1.5 %	1.2 %
Residues in disallowed regions	4.7%	4.5%	4.7%	4.5%	4.7%

After refinement, the model 2 showed 76.4% residues in the favorable region in the Ramachandran plot, with GDT-HA score 0.9015, RMSD value 0.538, MolProbity2.121, Clash score 8.4 and Poor rotamers score 0.5.

Model 2 was subjected to a second refinement to get better results. Five models resulted from the Galaxy refine server and they were evaluated depending on the Ramachandran plot statistical analysis using Prochek. The results shown in the Table (52):

Table (52) shows Ramachandran plot statistical analysis for the second refinement for candidate (A):

	Model 1	Model 2	Model 3	Model 4	Model 5
Residues in most favored regions	79.4%	79.4%	80.1%	79.1%	77.6%
Residues in additional allowed regions	15.2%	15.4%	13.9%	15.7%	16.9%
Residues in generously allowed regions	1.2%	1.2 %	2.0 %	1.2 %	1.5 %
Residues in disallowed regions	4.2%	4.0%	4.0%	4.0%	4.0%

After the second refinement, the model 3 showed 80.1% residues in the favorable region in the Ramachandran plot, with GDT-HA score 0.9732, RMSD value 0.345, MolProbity 1.919, Clash score 5.6 and Poor rotamers score 8. Ramachandran plots were created for before and after refinement models. Also, the before and after refinement models were inserted into ProSA-web protein structure analysis to obtain their Z-score. Results are shown in Figure (2).

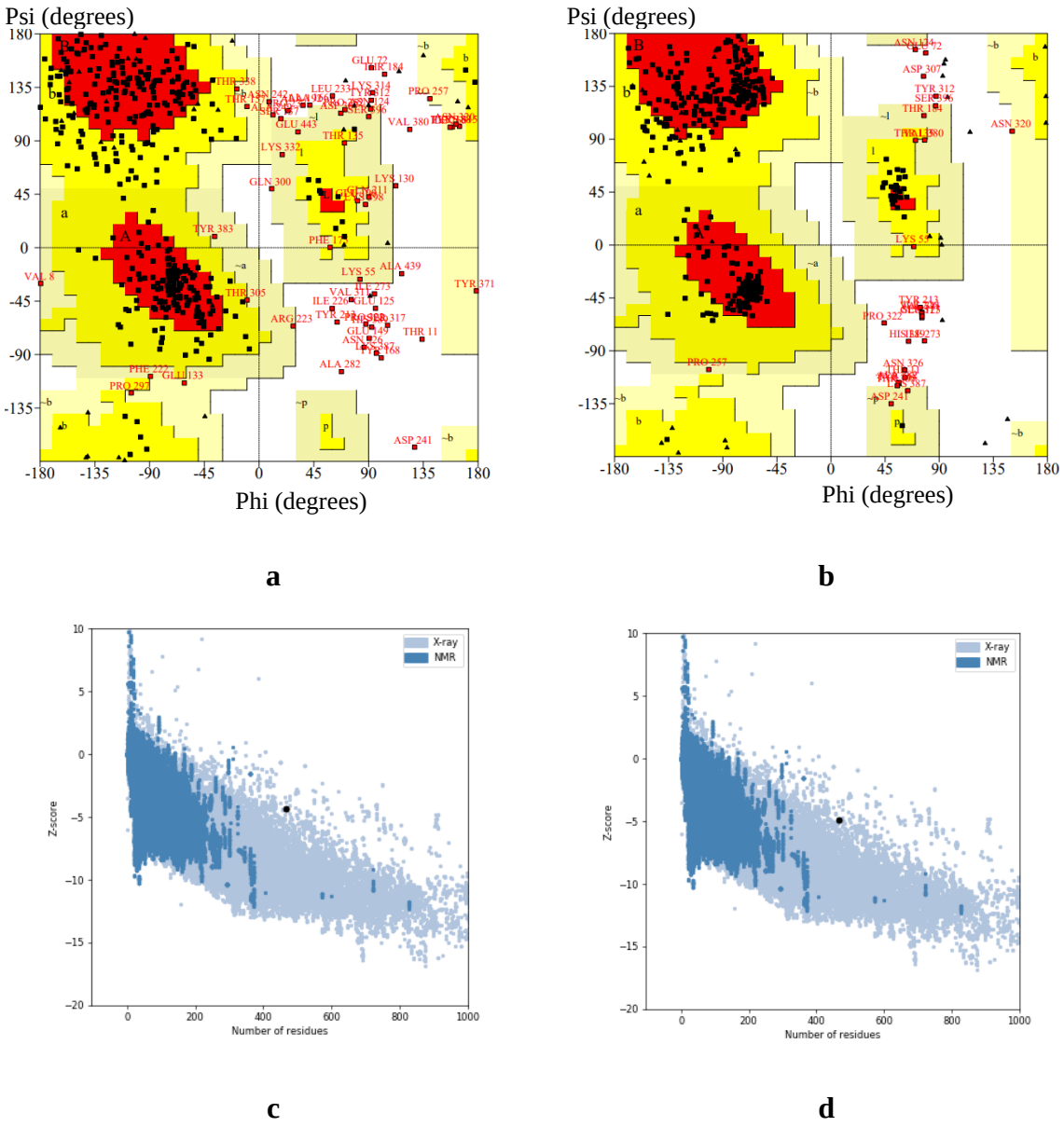


Fig (2)

a: Ramachandran plot for the A candidate before refinement shows:60.7% of residues in most favored regions, 26.4% residues in additional allowed regions, 6.7% residues in generously allowed regions and 6.2% residues in disallowed regions.

b: Ramachandran plot for the refined A shows: 80.1% of residues in most favored regions, 13.9% residues in additional allowed regions, 2.0 % residues in generously allowed regions and 4.0% residues in disallowed regions.

c: The z-score of the A candidate before refinement= -4.35 is slightly in range of native protein conformation. It is depicted in a large black spot. z-Score plot consists of z-scores of all experimentally protein chains in PDB defined by NMR spectroscopy

(dark blue) and X-ray crystallography (light blue).
d: The z-score of the A candidate before refinement= -4.9 is slightly in range of native protein conformation. It is depicted in a large black spot. z-Score plot consists of z-scores of all experimentally protein chains in PDB defined by NMR spectroscopy (dark blue) and X-ray crystallography (light blue).

4-7-2- Candidate (C) Refinement:

Candidate (C) refinement resulted in five models. The models were evaluated by Ramachandran plots statistical analysis. The results shown in the Table (53):

Table (53) shows Ramachandran plot statistical analysis for the refined models for candidate (C) by Galaxy refine web server:

	Model 1	Model 2	Model 3	Model 4	Model 5
Residues in most favoured regions	74.1%	75.5%	74.4%	76.0%	75.5%
Residues in additional allowed regions	18.5%	18.7%	19.6%	17.4%	17.6%
Residues in generously allowed regions	2.2 %	1.1 %	1.1 %	1.7%	1.9%
Residues in disallowed regions	5.2%	4.7%	5.0%	5.0%	5.0%

After refinement, the model 2 showed 75.5% residues in the favorable region in the Ramachandran plot, with GDT-HA score 0.9061, RMSD value 0.569, MolProbity2.501, Clash score 19.0 and Poor rotamers score 1.1. Ramachandran plots before and after refining and Z-score plots are shown in Figure (3).

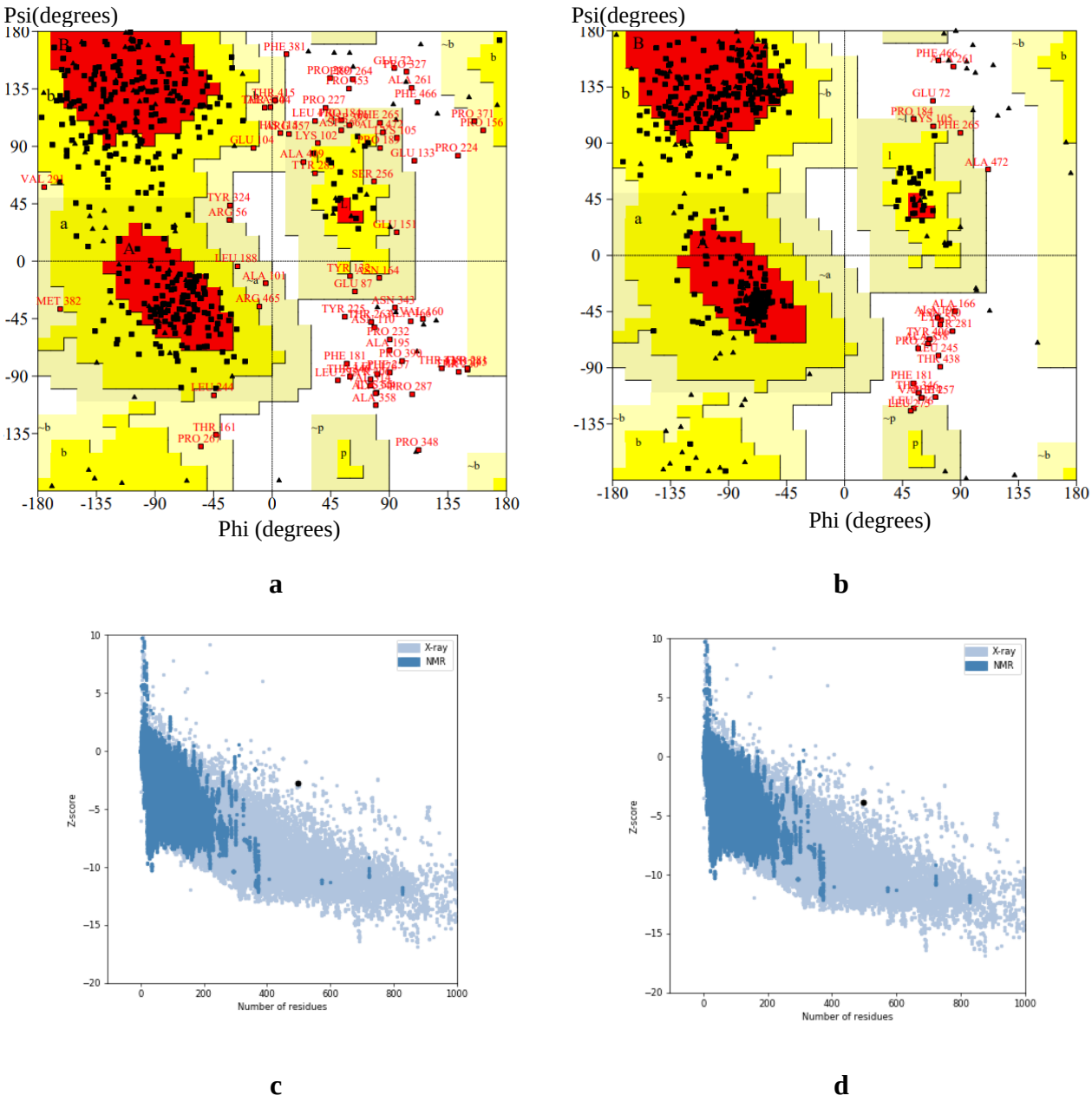


Fig (3)

a: Ramachandran plot for the C candidate before refinement shows:57.0% of residues in most favored regions, 27.8% residues in additional allowed regions, 8.3% residues in generously allowed regions and 6.9% residues in disallowed regions.

b: Ramachandran plot for the refined C shows: 75.5% of residues in most favored regions, 18.7% residues in additional allowed regions, 1.1 % residues in generously allowed regions and 4.7% residues in disallowed regions.

c: The z-score of the C candidate before refinement= -2.75 is slightly in range of native protein conformation. It is depicted in a large black spot. z-Score plot consists of z-scores of all

experimentally protein chains in PDB defined by NMR spectroscopy (dark blue) and X-ray crystallography (light blue).
d: The z-score of the C candidate before refinement= -3.86 is slightly in range of native protein conformation. It is depicted in a large black spot. z-Score plot consists of z-scores of all experimentally protein chains in PDB defined by NMR spectroscopy (dark blue) and X-ray crystallography (light blue).

4-8- Defining discontinuous B-cell epitopes (conformational):

The refined vaccine model for vaccine candidate (A) was subjected to B-cell epitopes prediction using Ellipro Server [68] in order to predict conformational B-cell epitopes. Discontinuous B-cell epitopes were predicted with scores ranging from 0.981 to 0.546.
Amino acid residues, the number of residues, sequence location and their scores have been listed in the table (54) below:

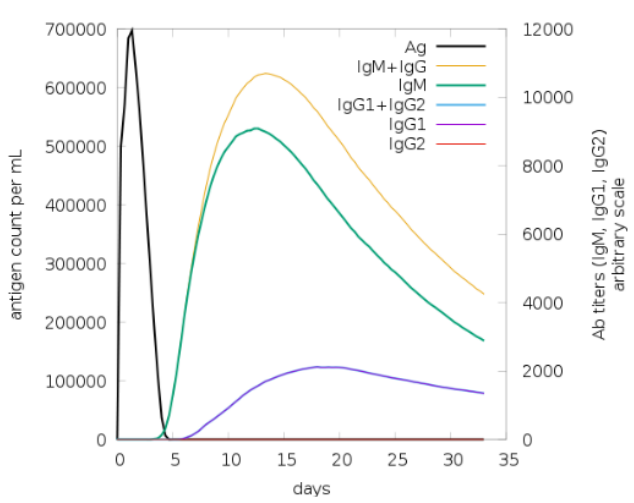
Table (54) shows the predicted amino acid residues, number of residues, sequence location and their scores for the discontinuous epitopes for the candidate (A):

No.	Residues	Number of residues	Score
1	_:D456, _:K459, _:K463	3	0.981
2	_:Y457, _:K458, _:K460, _:K461, _:K464, _:K465	6	0.967
3	_:N328, _:A329, _:A330, _:T331, _:K332, _:K333, _:N334, _:G335, _:T336, _:G337, _:T338, _:N339, _:A340, _:G341, _:L342, _:S343, _:D344, _:L345, _:K346, _:L347, _:D348, _:D349, _:K350, _:K351, _:D352, _:V353, _:E354, _:I355, _:N356, _:P357	30	0.801
4	_:A280, _:V281, _:A282, _:S283, _:S284, _:T285, _:S286, _:A287, _:S288, _:V289, _:A290, _:K291, _:A293	13	0.725
5	_:M1, _:G21, _:T22, _:P23, _:Q24, _:N25, _:T27, _:D28, _:L29, _:C30, _:A31, _:E32, _:Y33, _:N35, _:T36, _:Q37, _:I38, _:Y39, _:T40, _:L41, _:N42, _:D43, _:K44, _:I45, _:F46, _:S47, _:Y48, _:T49, _:E50, _:S51, _:L52, _:A53, _:G54, _:K55, _:R56, _:E57, _:M58, _:A59, _:I60, _:I61, _:T62, _:F63, _:K64, _:N65, _:G66, _:A67, _:I68, _:F69, _:Q70, _:V71, _:E72, _:V73, _:P74, _:G75, _:S76, _:Q77, _:H78, _:I79, _:D80, _:S81, _:Q82, _:K83, _:I86, _:K90, _:Y97, _:T99, _:E100, _:A101, _:K105, _:L106, _:C107, _:V108, _:W109, _:N110, _:N111, _:K112, _:T113, _:P114, _:H115, _:A116, _:I117, _:A118, _:A119, _:G379, _:V380, _:T381, _:Y382	87	0.699

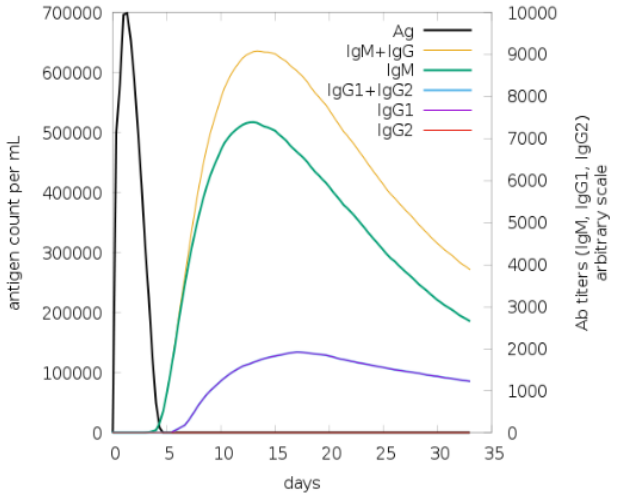
6	_:M169, _:L170, _:N171, _:D172, _:S173, _:W174, _:F175, _:L176, _:A177, _:A178, _:Y179, _:L180, _:P181, _:N182, _:Y183, _:P185, _:P186, _:A187, _:Y188, _:A189, _:A190, _:Y191, _:A192, _:T193, _:A194, _:D195, _:G196, _:T197, _:V198, _:V199, _:Y200, _:Q211, _:P212, _:Y213, _:G214, _:P215, _:G216, _:P217, _:G218, _:N221, _:F222, _:S224, _:Y225, _:I226, _:S227, _:Y228, _:F230, _:A415, _:A417, _:S418, _:S419, _:T420, _:S421, _:K422, _:R424, _:Q425, _:L426, _:E427, _:Q428, _:T429, _:K430, _:Q431, _:R432, _:F433, _:E434, _:V435, _:G436, _:L437, _:S438, _:A439, _:K440, _:K441, _:T442, _:E443, _:L444, _:R445, _:L446, _:D447, _:I448, _:R449, _:K450, _:M451, _:D452, _:K453, _:M454, _:R455	86	0.66
7	_:R138, _:L139, _:K140, _:A141, _:A142, _:Y143, _:S145, _:D146, _:N147, _:G148, _:E149, _:D150, _:P295, _:G296, _:P297, _:G298, _:G299, _:Q300, _:A301, _:V302, _:V303, _:A304, _:T305, _:A306, _:D307, _:G308, _:T309, _:V310, _:V311	29	0.546

4-9-Immune response simulation:

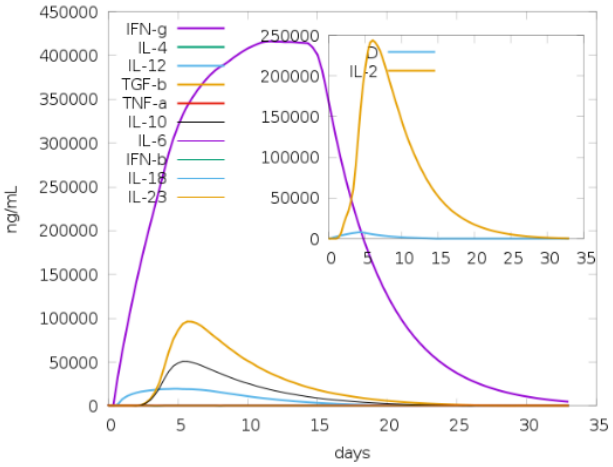
C-ImmSim simulates the three main components of the functional mammal system (Thymus, lymph node, and bone marrow). The input parameters for the immune simulations were as follow: volume (10), HLA (A0101, A0301, B3501, DRB1_0701, DRB1_0301), random seed (12345), number of steps (100), number of injections set to 1. The rest of the parameters were considered to be the default. Immune response simulation is shown in detail in Figure (4).



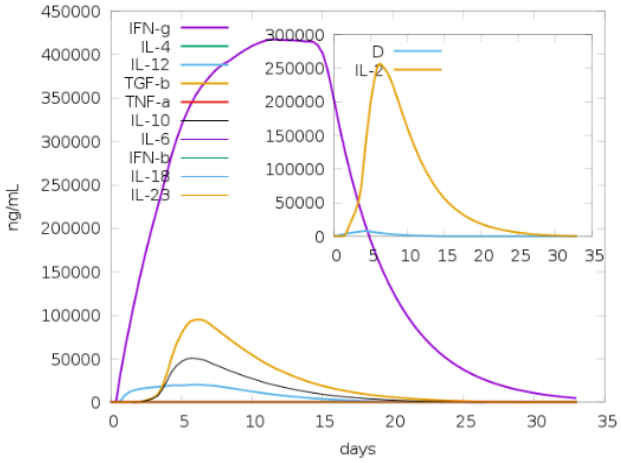
a



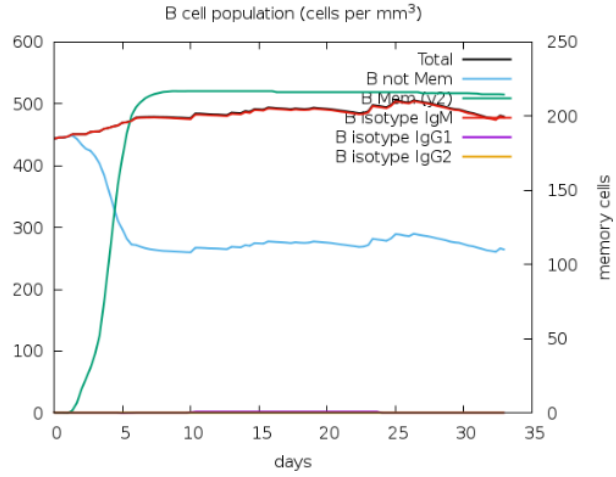
b


IFN-g
IL-4
IL-12
TGF-b
TNF-a
IL-10
IL-6
IFN-b
IL-18
IL-23
D
IL-2
ng/mL
days

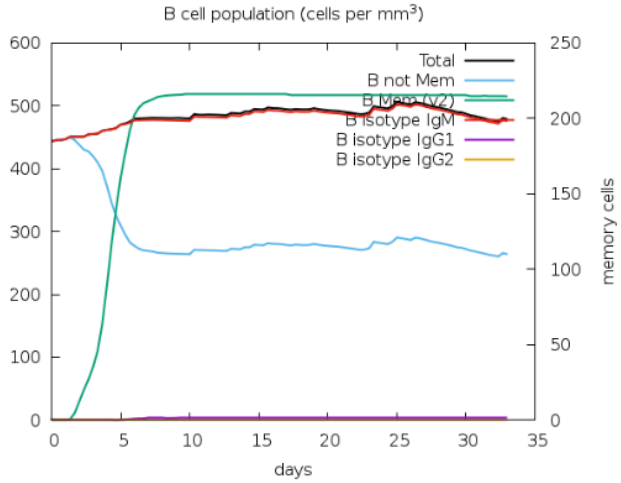
c


IFN-g
IL-4
IL-12
TGF-b
TNF-a
IL-10
IL-6
IFN-b
IL-18
IL-23
D
IL-2
ng/mL
days

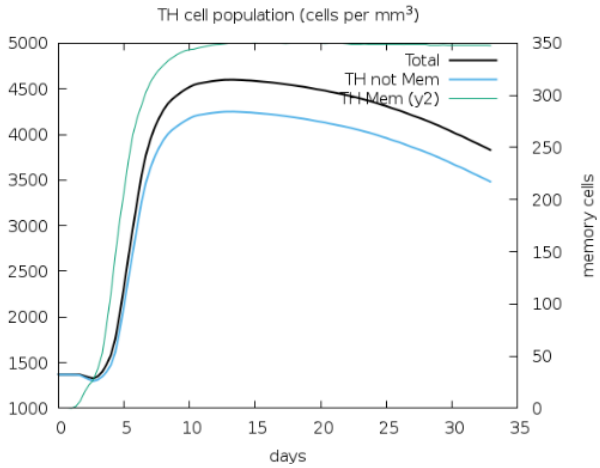
d


B cell population (cells per mm3)
Total
B not Mem
B Mem (y2)
B isotype IgM
B isotype IgG1
B isotype IgG2
memory cells
days

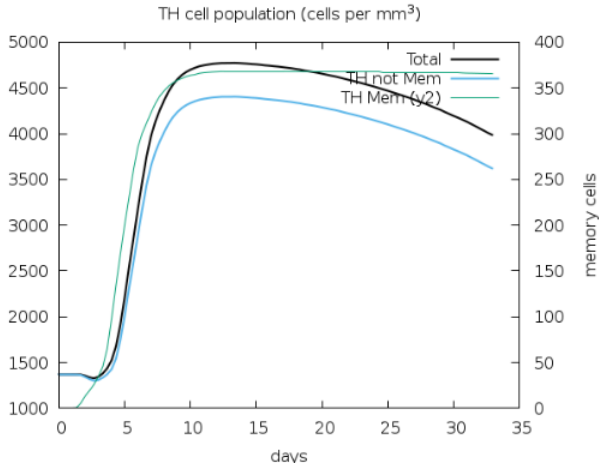
e


B cell population (cells per mm3)
Total
B not Mem
B Mem (y2)
B isotype IgM
B isotype IgG1
B isotype IgG2
memory cells
days

f


TH cell population (cells per mm3)
Total
TH not Mem
TH Mem (y2)
memory cells
days

g


TH cell population (cells per mm3)
Total
TH not Mem
TH Mem (y2)
memory cells
days

h

Fig (4)

a and b show Antigen and immunoglobulin responses to the antigen after injection for vaccine candidate (A) and vaccine candidate (C) respectively.
c and d show Cytokines. Concentration of cytokines and interleukins for vaccine candidate (A) and vaccine candidate (C) respectively. D in the inset plot is a danger signal.
e and f show B cell population. B lymphocytes: total count, memory cells, and subdivided in isotypes IgM, IgG1 and IgG2 for vaccine candidate (A) and vaccine candidate (c) respectively.
g and h show TH cell population. CD4 T-helper lymphocytes count (total and memory counts) for vaccine candidate (A) and vaccine candidate (C) respectively.

In both candidates, the response is characterized by high IgG + IgM and IgM concentration, followed by IgG1 elevation with concomitant antigen reduction. Additionally, robust interleukin and cytokine response was observed. The IFN-gamma concentration was significantly high. Additionally, B-cell and T-helper populations were also increased with the injection. All of this indicates that both candidates made a successful immune response and clearance after subsequent encounters. These findings confirmed the immunogenicity of both candidates. The remaining procedures was conducted only on candidate (A).

4-10- Molecular docking:

4-10-1- Vaccine candidate (Ligand) and Toll-like receptor-2 (TLR2) (Receptor) docking:

The B pentamers of the Escherichia coli heat-labile enterotoxins interact functionally with Toll-like receptor-2 (TLR2) [76]. According to the high similarity between B pentamers of Cholera toxin (CTB) and the B pentamers of the Escherichia coli heat-labile enterotoxins in sequence and structure, TLR2 was assumed by us as an acceptable receptor in our study [14].

The 3D structure of the receptor (TLR2) was modeled by the Phyre2 server after the amino acid sequence was retrieved from the Uniprot (Uniprot ID: O60603). The modeled structure is shown in Figure (5).

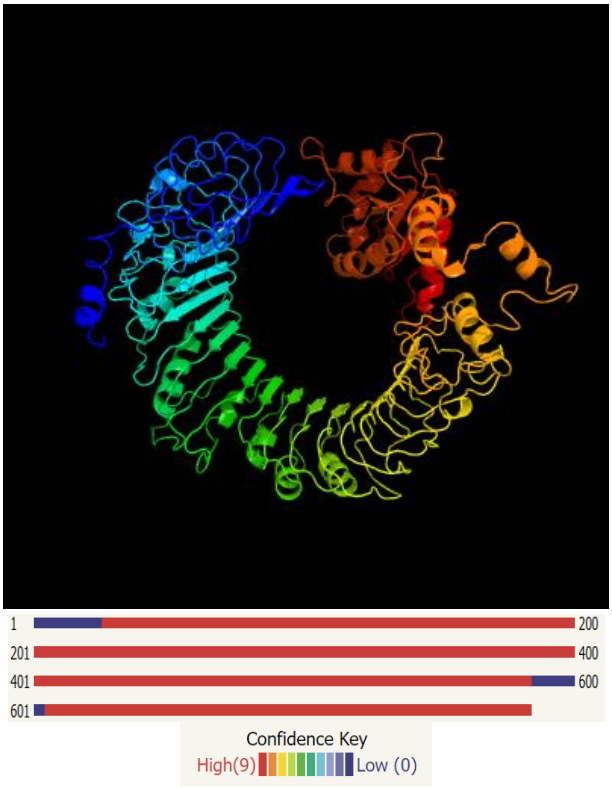


Fig (5)

3D model of the TLR2 modeled by the Phyre2 server. 94% of residues modeled at >90% confidence. The confidence of the generated model shows the high confidence of the model.

Patchdock [70] [71] was used for docking. The (vaccine candidate (A)-TLR2) top 20 models (solutions) were received via email with their respective docking score. These models were refined in Firedock [77] [78]. The best solution was solution 2, where the global Energy was 2.08, van der Waals energy (vdW) was-1.09, repulsive energy was 0.00, atomic contact energy (ACE) was 1.97, and hydrogen bond energy was 0.00. Solution 2 was visualized using the Discovery Studio 2021 program as it is shown in Figure (6).

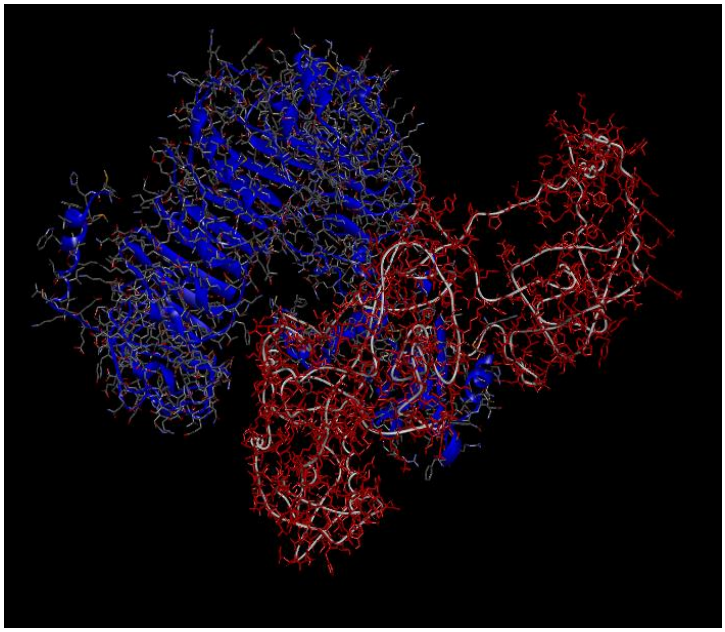


Fig (6)

The protein-protein interaction between vaccine candidate (Ligand in the red color) and TLR2 (Receptor in the dark blue color) visualized by Discovery Studio 2021 program.

4-10-2- Vaccine candidate (Ligand) and Toll-like receptor-4 TLR4 (Receptor) docking:

Another probable receptor for the Non-toxic B pentamers of Cholera toxin (CTX) (the adjuvant) is Toll-like receptor (TLR4) as it is responsible for immune response against 17D vaccine [73][79].

The 3D structure of the receptor (TLR4) was modeled by the Phyre2 server after the amino acid sequence was retrieved from the Uniprot (Uniprot ID: O00206). The modeled structure is shown in Figure (7).

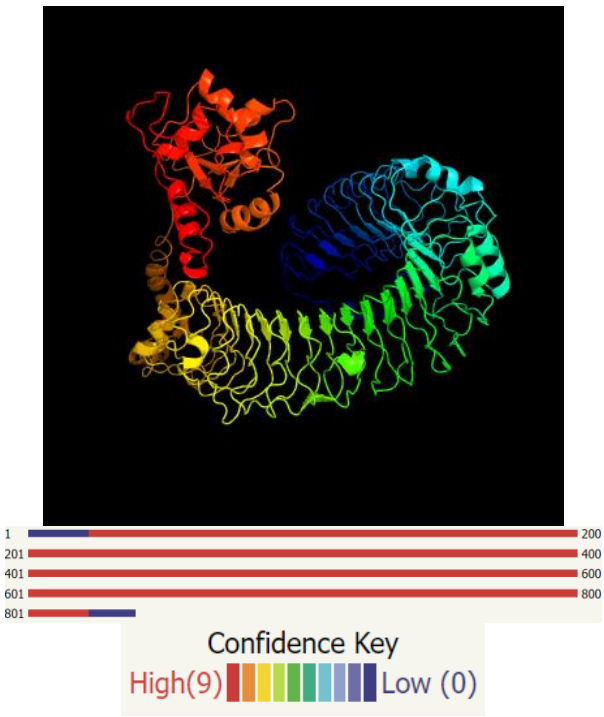


Fig (7)

3D model of the TLR2 modeled by the Phyre2 server. 95% of residues modeled at >90% confidence. The confidence of the generated model shows the high confidence of the model.

The docking between the vaccine candidate (A) (ligand) and TLR4 (receptor) was performed using Patchdock [70] [71]. The top 20 docked models (solutions) were received via email with their respective docking score. The top models were refined in Firedock [77] [78].The best solution was solution 5, where the global Energy was 15.78, van der Waals energy (vdW) was-4.71, repulsive energy was 2.32, atomic contact energy (ACE) was 3.13, and hydrogen bond energy was 0.00. Solution 5 was visualized using the Discovery Studio 2021 program as it is shown in Figure (8).

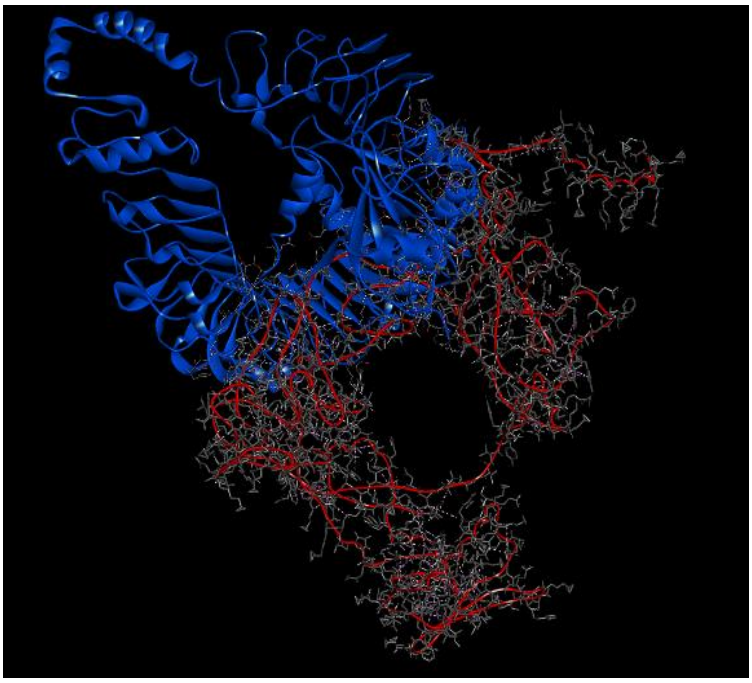


Fig (8)
The protein-protein interaction between vaccine candidate (Ligand in the red color) and TLR4 (Receptor in the dark blue color) visualized by Discovery Studio 2021 program.

4-11- Codon adaptation and in-silico cloning:

Codon optimization was carried out using Java Codon Adaptation Tool (JCat) server in order to maximize the production of the designed vaccine candidate in an appropriate expression system [80]. This step is necessary because the degeneracy of the genetic code allows most amino acids to be encoded by multiple codons. The coding sequence of the vaccine candidate was codon-optimized for protein expression in the E. coli (strain K12) host. JCat have a possibility to avoid restriction enzyme binding sites in the adapted DNA.

ATGATCAAACTGAAATTCGGTGTTTTCTTCACCGTTCTGCTGTCTTCTGC 50
TTACGCTCACGGTACCCCGCAGAACATCACCGACCTGTGCGCTGAATACC 100
ACAACACCCAGATCTACACCCTGAACGACAAAATCTTCTCTTACACCGAA 150
TCTCTGGCTGGTAAACGTGAAATGGCTATCATCACCTTCAAAAACGGTGC 200
TATCTTCCAGGTTGAAGTTCCGGGTTCTCAGCACATCGACTCTCAGAAAA 250
AAGCTATCGAACGTATGAAAGACACCCTGCGTATCGCTTACCTGACCGAA 300
GCTAAAGTTGAAAAACTGTGCGTTTGGAACAACAAAACCCCGCACGCTAT 350
CGCTGCTATCTCTATGGCTAACGAAGCTGCTGCTAAAAAAGTTTACGAAG 400
GTACCCTGTCTCGTCTGAAAGCTGCTTACTACTCTGACAACGGTGAAGAC 450
GGTTACGCTGCTTACGTTACCGAAACCAACGCTGCTAAATACGCTGCTTA 500
CTACATGCTGAACGACTCTTGGTTCCTGGCTGCTTACCTGCCGAACTACA 550
CCCCGCCGGCTTACGCTGCTTACGCTACCGCTGACGGTACCGTTGTTTAC 600
GCTGCTTACCGTTCTGACCCGACCATCACCCAGCCGTACGGTCCGGGTCC 650

GGGTAAACCGAACTTCCGTTCTTACATCTCTTACCAGTTCAACCTGCTGG 700
GTCCGGGTCCGGGTTTCACCGACAACATCTCTTTCGAAGTTCTGGCTGCT 750
ACCCCGTTCGGTCCGGGTCCGGGTGAAGCTAAACAGCTGGCTCTGATCTC 800
TGTTTACTCTTACCCGATCGGTCCGGGTCCGGGTGTTGCTGTTGCTTCTT 850
CTACCTCTGCTTCTGTTGCTAAAGCTGCTGGTCCGGGTCCGGGTGGTCAG 900
GCTGTTGTTGCTACCGCTGACGGTACCGTTGTTTACTCTAAAAAACGTTC 950
TCTGGGTAACTACCCGGCTACCGCTAACGCTAACGCTGCTACCAAAAAAA 1000
ACGGTACCGGTACCAACGCTGGTCTGTCTGACCTGAAACTGGACGACAAA 1050
AAAGACGTTGAAATCAACCCGTGGGTTTTCATGATCGCTGGTGGTTACAA 1100
AAAAAAAGAATACACCATCGGTGACGCTCTGATCGGTGTTACCTACTACA 1150
ACGCTAAAAAAACCCACACCCAGTACGAAGCTTACTCTGGTAAATACGAA 1200
GAACTGGAAAAAAAAGGTGGTACCGGTGGTGCTGCTACCGTTGCTGTTGC 1250
TTCTTCTACCTCTAAAAAACGTCAGCTGGAACAGACCAAACAGCGTTTCG 1300
AAGTTGGTCTGTCTGCTAAAAAAACCGAACTGCGTCTGGACATCCGTAAA 1350
ATGGACAAAATGCGTGACTACAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1400
A

The output of the (JCat) server included two parameters: codon adaptation index (CAI) was 1.0 (ideal score between 0.8 and 1) and the percentage of GC content was 49.817651349380014 (should be between 30 and 70%), indicated that the expression of the vaccine candidate (A) was efficient and potentially stable in the E. coli K-12 strain. The candidate codons lacked restriction sites for XhoI and NcoI which were added manually. The AAA codons in the (3') end were replaced by AAG as both codons encode lysine.

CATGGGAATCAAACTGAAATTCGGTGTTTTCTTCACCGTTCTGCTGTCTTCTGCTTACGCTCACGGTACCCCGCAGA
ACATCACCGACCTGTGCGCTGAATACCACAACACCCAGATCTACACCCTGAACGACAAAATCTTCTCTTACACCGA
ATCTCTGGCTGGTAAACGTGAAATGGCTATCATCACCTTCAAAAACGGTGCTATCTTCCAGGTTGAAGTTCCGGGT
TCTCAGCACATCGACTCTCAGAAAAAAGCTATCGAACGTATGAAAGACACCCTGCGTATCGCTTACCTGACCGAA
GCTAAAGTTGAAAAACTGTGCGTTTGGAACAACAAAACCCCGCACGCTATCGCTGCTATCTCTATGGCTAACGAAG
CTGCTGCTAAAAAAGTTTACGAAGGTACCCTGTCTCGTCTGAAAGCTGCTTACTACTCTGACAACGGTGAAGACGG
TTACGCTGCTTACGTTACCGAAACCAACGCTGCTAAATACGCTGCTTACTACATGCTGAACGACTCTTGGTTCCTGG
CTGCTTACCTGCCGAACTACACCCCGCCGGCTTACGCTGCTTACGCTACCGCTGACGGTACCGTTGTTTACGCTGCT
TACCGTTCTGACCCGACCATCACCCAGCCGTACGGTCCGGGTCCGGGTAAACCGAACTTCCGTTCTTACATCTCTTA
CCAGTTCAACCTGCTGGGTCCGGGTCCGGGTTTCACCGACAACATCTCTTTCGAAGTTCTGGCTGCTACCCCGTTCG
GTCCGGGTCCGGGTGAAGCTAAACAGCTGGCTCTGATCTCTGTTTACTCTTACCCGATCGGTCCGGGTCCGGGTGTT
GCTGTTGCTTCTTCTACCTCTGCTTCTGTTGCTAAAGCTGCTGGTCCGGGTCCGGGTGGTCAGGCTGTTGTTGCTAC
CGCTGACGGTACCGTTGTTTACTCTAAAAAACGTTCTCTGGGTAACTACCCGGCTACCGCTAACGCTAACGCTGCT
ACCAAAAAAAACGGTACCGGTACCAACGCTGGTCTGTCTGACCTGAAACTGGACGACAAAAAAGACGTTGAAATC
AACCCGTGGGTTTTCATGATCGCTGGTGGTTACAAAAAAAAAGAATACACCATCGGTGACGCTCTGATCGGTGTTA
CCTACTACAACGCTAAAAAAACCCACACCCAGTACGAAGCTTACTCTGGTAAATACGAAGAACTGGAAAAAAAAG
GTGGTACCGGTGGTGCTGCTACCGTTGCTGTTGCTTCTTCTACCTCTAAAAAACGTCAGCTGGAACAGACCAAACA
GCGTTTCGAAGTTGGTCTGTCTGCTAAAAAAACCGAACTGCGTCTGGACATCCGTAAAATGGACAAAATGCGTGAC
TACAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGCTCGA

The vaccine candidate (A) was cloned and expressed in E. coli using pET23d vector by Geneious program as it is shown in Figure (9).

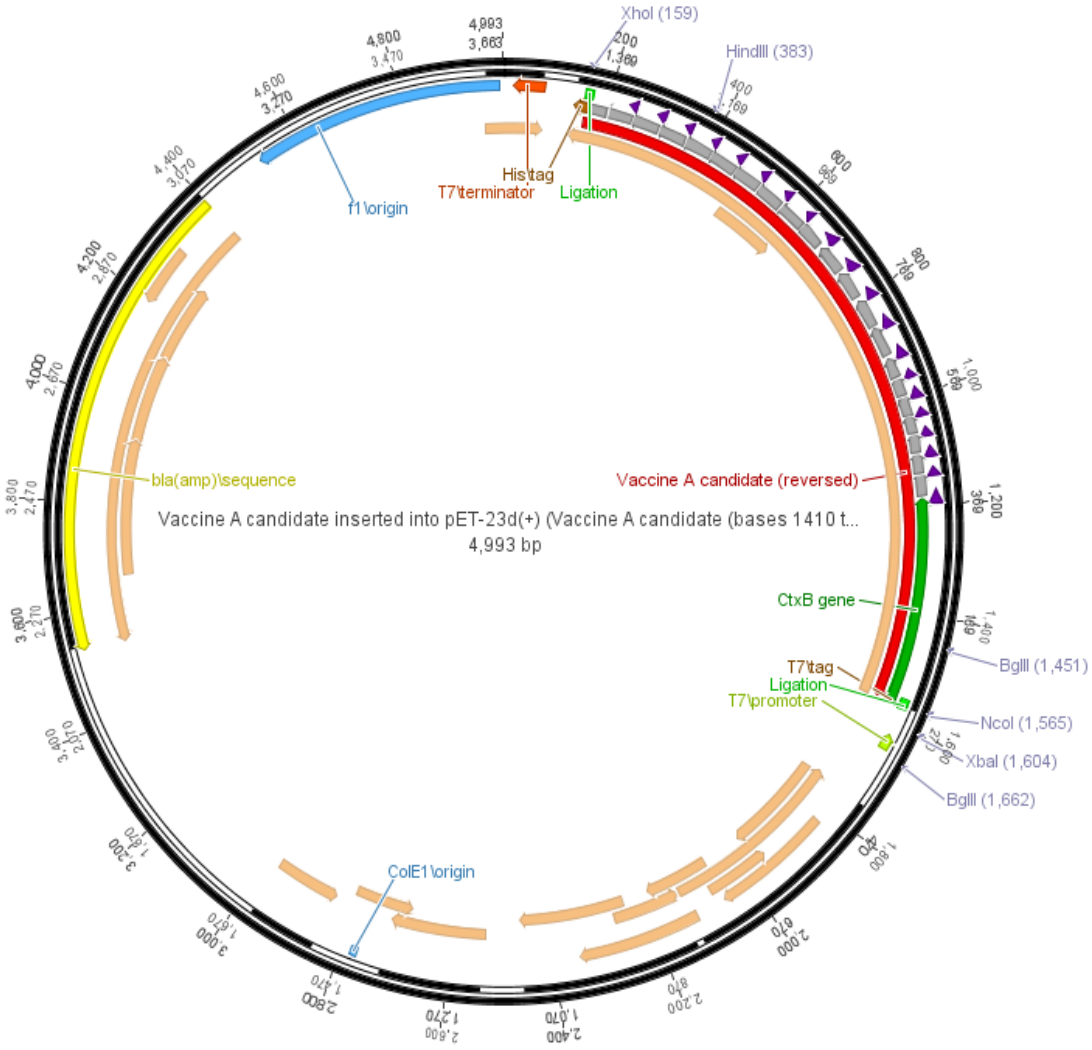


Fig (9)

shows In-silico cloning of the vaccine candidate into pET23d (+) expression vector using Geneious software. The red part represents the coding sequence of the vaccine candidate, and the black circle represents the vector backbone.

The vaccine candidate (A) sequence can be synthesized by a company that provides a service of DNA synthesis such as Synplogen. Synplogen uses one-step DNA assembly to create designer DNA [81]

5- Discussion:

Cholera is an acute watery diarrheal infection that can lead to death if left untreated, it is caused by ingestion of food or water contaminated with the bacterium *Vibrio cholerae (V. cholerae)* [1]. Currently there are three World Health Organization (WHO) pre-qualified oral cholera vaccines (OCV): Dukoral®, Shanchol™, and Euvichol®[3]. But they have some limitations. Thus, it was suggested to develop a new generation of vaccines using immunoinformatics. One of the main applications of immunoinformatics is recognizing immunoprotective antigens and developing B and T cell epitope prediction algorithms which decrease the time and costs required for experimental analysis [7]. The use of immunoinformatics has been accelerated toward the design of a multiepitope vaccine candidate because of their various advantages, which include high specificity, good safety, stability, and easy production and storage. Antibodies to *V. cholerae* LPS mediate protection against cholera. But also the existence of non-LPS protective antigens has also been documented to play an important role in protection via inhibition of intestinal colonization of vibrio [26].

In this study, a robust immunoinformatics approach was used to design a novel multi-epitope vaccine against *V. Cholerae*. *V. cholerae* O1 biovar El tor str. N16961 was selected as a reference strain. According to our survey, the most literary supported and frequently mentioned proteins in the medical literature were nominated and their amino acid sequences were retrieved in Fasta format to be subjected to virulent, antigenicity and subcellular localization predictions through VFDB, Vaxijen 2.0 and psortb respectively.

Epitopes were predicted using bioinformatics tools (considering the classification of Human Leukocyte Antigen (HLA) Supertypes in the Syrian population) through MHCI prediction tool Immune Epitope Database Analysis Resource for Cytotoxic T Lymphocyte (CTL) epitopes prediction and MHC II prediction tool from Immune Epitope Database Analysis Resource for Helper T Lymphocyte (HTL) Epitopes Prediction and ABCpred for B Lymphocyte Epitopes Prediction. The sequence of each predicted epitope was entered into Vaxijen 2.0 server in order to evaluate the capacity to prompt an immune response, AllerTOP v. 2. for allergenicity prediction, ToxinPred to determine the toxicity. Class I Immunogenicity from IEDB Analysis Resource was used in order to compute immunogenicity score for CTL epitopes and I IL4pred platform and IFN-γ epitope server to predict whether the HTL epitopes have ability to induce IL4 and IFN-γ or not. As a result, Seven CTL, five HTL and eight LBL epitopes were accepted to be included in the final vaccine design.

The issue of population coverage in relation to MHC polymorphism is further complicated by the fact that different HLA types are expressed at dramatically different frequencies in different ethnicities. In regard to MHC I polymorphism, the Syrian population is genetically closer to neighboring human populations, (Jordanians, Lebanese, and Turks) and to Europeans in the north of the Mediterranean. The Syrians are genetically far from human populations from the Arabian Peninsula and North Africa, and very far from the Chinese and other African human populations. In regard to MHC II polymorphism, the population of Syria is genetically closer to Lebanese,

Jordanians, and Iranians compared with Austrians, Italians, and Koreans. The greater genetic distances from our population were seen in Chinese

Several vaccine candidates were proposed based on different linkers between epitopes and different adjuvant sites. The physicochemical properties of the proposed vaccine candidates, solubility, antigenicity, allergenicity, homology with homo-sapiens proteins and toxicity properties were assessed using various tools. The best candidate depending on its physicochemical properties was chosen. Another good candidate was chosen for comparison.

The best-chosen candidate was the one which was constructed by connecting the seven CTL, five HTL, and eight LBL epitopes using the AAY, GPGPG and KK linkers, respectively. CTB adjuvant enhances antigen immune response, vaccine stability, and longevity. It was added to the protein sequence in the N terminal using the EAAAK linker. The vaccine candidate (A) was of length of 457 amino acids and a molecular weight of 49.395 kDa which were good length and molecular weight compared with other studies. The vaccine candidate had a theoretical PI of 9.13, which indicated that the vaccine was basic and protein provides a stable connection in the physiological pH range. An aliphatic index of 72.24 (>70) suggests that a vaccine is thermostable over a wide temperature range. A negative grand average hydropathy value indicates that the vaccine is hydrophilic, which has a good contribution to a water-based environment. Furthermore, the antigenicity, allergenicity, non-human homology and toxicity results revealed that the vaccine candidate was antigenic, non-allergenic, have no homology with human proteins and nontoxic. The vaccine was insoluble and a solubility-enhancing peptide tag (10 Lysine residues) was added in order to enhance solubility. Physicochemical properties, antigenicity, allergenicity, homology with homo-sapiens proteins and toxicity properties of the vaccine candidate were reassessed after adding solubility-enhancing peptide tag and no significance changes were obtained. The vaccine candidate after adding the solubility enhancing tag was of length of 467 amino acids, a molecular weight of 50.677 KD, theoretical PI of PI was 9.45, Aliphatic index of 71.18 and grand average of hydropathicity of -0.410. The solubility was enhanced.

The secondary structure analysis revealed that the vaccine consists of 34.90% a-helix, 24.84% b-strand and 40.26% random coils. This structure has high hydrogen bond energy that enables good interactions with antibodies. The 3D structure was predicted using the Phyre2 server. Ramachandran plot was created for the predicted model. A good quality model is expected to have over 90% in the most favored. As a result, the vaccine needs further double refinement by Galaxy server which helps rebuild and repackage side chains in the model in order to enhance the overall quality. 5 models were obtained after refinement and Ramachandran plots were used to assess the best one. Model 3 was the best model in the second refinement. Ramachandran plots for the models before and after refinement were created. Also, the before and after refinement models were inserted into ProSA-web protein structure analysis to check 3D models of protein structures for potential errors. The refined 3D protein vaccine model was subjected to B-cell epitopes prediction using Ellipro Server in order to predict conformational B-cell epitopes as B-cell epitopes play an important role in humoral response.

The immune simulation assessment revealed a high IgG + IgM and IgM concentration, followed by IgG1 elevation with concomitant antigen reduction. Additionally, robust interleukin and cytokine response was observed. The IFN-gamma concentration was significantly high. Additionally, B-cell and T-helper populations were also increased with the injection. All of this indicates that the vaccine induces a successful immune response and clearance after subsequent encounters.

Molecular docking was used to determine the interactions between ligand and receptor molecules. TLR2 and TLR4 were used as receptors to dock the vaccine candidate. The receptors were modeled using the Phyre2 server after their amino acid sequences were retrieved from the uniprot. Patchdock was used for the docking. Firedock refinement after Patchdock showed that (vaccine candidate (A) and TLR2) best solution is solution 2, where the global energy was 2.08, van der Waals energy (vdW) was-1.09, repulsive energy was 0.00, atomic contact energy (ACE) was 1.97, and hydrogen bond energy was 0.00. And (vaccine candidate (A) and TLR4) best solution was solution 5, where the global Energy was 15.78, van der Waals energy (vdW) was-4.71, repulsive energy was 2.32, atomic contact energy (ACE) was 3.13, and hydrogen bond energy was 0.00. Solution 5 was visualized using the Discovery Studio 2021 program. Discovery Studio 2021 program was used to visualize results.

In order to express the gene of interest in a vector, the Java Codon Adaptation Tool (JCat) was used, which corroborates codon compliance by optimizing the vaccine candidate sequence. The designed candidate was optimized according to E. coli strain-K12. The vaccine candidate was cloned in E. coli using pET23d vector by Geneious program.

The current study was based on integrated computational tools and it lacks in-vivo and in-vitro evaluation. That was a limitation of this study. The protective efficacy and the safety of the designed vaccine should be validated through further experimental assessment.

6- Conclusions:

A potential multi epitope vaccine against *Vibrio cholerae* was designed using various immunoinformatics tools which was a strength factor of this study beside using up to date immunological information. The vaccine was designed based on CTL, HTL and LBL epitopes which were linked using linkers. N terminal adjuvant was added to improve antigenicity. The vaccine had satisfied antigenicity, physicochemical properties, allergenicity, toxicity, and immunogenicity properties. But unsatisfied solubility score which imposed adding solubility enhancing tag. The added tag didn't significantly change the properties. The immune simulation measures showed their potential to trigger and induce immune response. Future study is required to evaluate efficacy and the safety of the designed vaccine because the current study lacks in-vivo and in-vitro evaluation although it is based on integrated computational tools. That was a limitation of this study.

Study tools are shown in the Table (55):

Table (55) shows the tools which were used in the study with a brief descriptions and their references:

No.	Tool	Description	Reference
1	VFDB	Integrated and comprehensive online resource for curating information about virulence factors of bacterial pathogens.	[22,23]
2	Vaxijen 2.0	A server for prediction of protective antigens, tumor antigens and subunit vaccines.	[24]
3	psortb,	The most precise bacterial protein subcellular localization (SCL) predictor since it was first made available in 2003	[25]
4-5-6-7-8	IEDB Analysis Resource	A website which provides a collection of tools for the prediction and analysis of immune epitopes. It serves as a companion site to the Immune Epitope Database (IEDB), a manually curated database of experimentally characterized immune epitopes. It contains MHC I and II epitopes prediction, B cell epitopes prediction, Class I Immunogenicity prediction and population coverage of T-cell epitope prediction.	[37–39] [45] [51] [41] [68]
9	AllerTOP v. 2.0	Bioinformatics tool for allergenicity prediction	[42]
10	ToxinPred	A unique in-silico method of its kind, which was useful in predicting toxicity of peptides/proteins. It was useful in designing least toxic peptides and discovering toxic regions in proteins	[44]
11	IFN-γ epitope server	IFN-γ prediction for HTL epitopes by (svm based method).	[47]

12	IL4pred	Insilico platform for designing and discovering of Interleukin-4 inducing peptides. IL4pred allows users to predict whether their peptides have the ability to induce IL4 or not, in simple words it allows them to predict IL4 inducing peptides or IL4 peptides.	[48]
13	ABCpred	This server assist in locating epitope regions that are useful in selecting synthetic vaccine candidates using an artificial neural network.	[49,50]
14	ProtParam tool / Expasy	ProtParam is a tool which allows the computation of various physical and chemical parameters for a given protein stored in Swiss-Prot or TrEMBL or for a user entered protein sequence. The computed parameters include the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY).	[57]
15	Protein-Sol	Protein-sol is a simple and free, web based suite of theoretical calculations and predictive algorithms for understanding protein solubility and stability. The scaled solubility value (QuerySol) is the predicted solubility. The population average for the experimental dataset (PopAvrSol) is 0.45, and therefore any scaled solubility value greater than 0.45 is predicted to have a higher solubility than the average soluble E.coli protein from the experimental solubility dataset, and any protein with a lower scaled solubility value is predicted to be less soluble.	[58]
16	SOLpro predictor	SOLpro predicts the propensity of a protein to be soluble upon overexpression in E. coli using a two-stage SVM architecture based on multiple representations of the primary sequence.	[59]

17	SOPMA server	Predicts the secondary structural features include a-helix, b-strand and random coils that were evaluated	[62]
18	Phyre2 server	Homology modeling	[63]
19	Galaxy refine	Helps rebuild and repackage side chains in the 3D mode.	[75]
20	Prochek	Generating Ramachandran plots.	[64]
21	ProSA-web protein structure analysis	ProSA is a tool widely used to check 3D models of protein structures for potential errors. The z-score indicates overall model quality and measures the deviation of the ttotal energy of the structure with respect to an energy distribution derived from random conformations. Z-scores outside a range characteristic for native proteins indicate erroneous structures.	[65] [65,66]
22	Java Codon Adaptation Tool (JCat) .	The CodonAdaptationTool (JCAT) presents a simple method to adapt the Codon Usage to most sequenced prokaryotic organisms and selected eukaryotic organisms. The codon adaptation plays a major role in cases where foreign genes are expressed in hosts and the codon usage of the host differs from that of the organism where the gene stems from. Unadapted codons in the host can for example lead to a minor expression rate.	[69]
23	Patchdock	Molecular Docking Algorithm Based on Shape Complementarity Principles	[70][71]
24	Firedock	Patchdock results refining server	[77][78]
25	Geneious software	One platform with all the molecular biology and sequence analysis tools organization needs	
26	Discovery studio software	Viewing, sharing, and analyzing protein and modeling data.	

References:

1. Cholera. [cited 19 Oct 2022]. Available: https://www.who.int/news-room/fact-sheets/detail/cholera
2. Shaikh H, Lynch J, Kim J, Excler J-L. Current and future cholera vaccines. Vaccine. 2020;38 Suppl 1: A118–A126. doi:10.1016/j.vaccine.2019.12.011
3. Immunization, Vaccines and Biologicals. [cited 19 Oct 2022]. Available: https://www.who.int/teams/immunization-vaccines-and-biologicals/diseases/cholera
4. Tomar N, De RK. Immunoinformatics: an integrated scenario. Immunology. 2010;131: 153–168. doi:10.1111/j.1365-2567.2010.03330.x
5. Sit B, Fakoya B, Waldor MK. Emerging concepts in cholera vaccine design. Annu Rev Microbiol. 2022;76: 681–702. doi:10.1146/annurev-micro-041320-033201
6. Yurina V, Adianingsih OR. Predicting epitopes for vaccine development using bioinformatics tools. Therapeutic Advances in Vaccines and Immunotherapy. 2022;10: 25151355221100216. doi:10.1177/25151355221100218
7. 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: 285–307. doi:10.1556/AMicr.61.2014.3.4
8. Seib KL, Zhao X, Rappuoli R. Developing vaccines in the era of genomics: a decade of reverse vaccinology. Clin Microbiol Infect. 2012;18 Suppl 5: 109–116. doi:10.1111/j.1469-0691.2012.03939.x
9. Flower DR, Macdonald IK, Ramakrishnan K, Davies MN, Doytchinova IA. Computer aided selection of candidate vaccine antigens. Immunome Res. 2010;6 Suppl 2: S1. doi:10.1186/1745-7580-6-S2-S1
10. Oli AN, Obialor WO, Ifeanyichukwu MO, Odimegwu DC, Okoyeh JN, Emechebe GO, et al. Immunoinformatics and vaccine development: an overview. Immunotargets Ther. 2020;9: 13–30. doi:10.2147/ITT.S241064
11. Pandey RK, Ojha R, Aathmanathan VS, Krishnan M, Prajapati VK. Immunoinformatics approaches to design a novel multi-epitope subunit vaccine against HIV infection. Vaccine. 2018;36: 2262–2272. doi:10.1016/j.vaccine.2018.03.042
12. Saha S, Vashishtha S, Kundu B, Ghosh M. In-silico design of an immunoinformatics based multi-epitope vaccine against Leishmania donovani. BMC Bioinformatics. 2022;23: 319. doi:10.1186/s12859-022-04816-6
13. Hajighahramani N, Nezafat N, Eslami M, Negahdaripour M, Rahmatabadi SS, Ghasemi Y. Immunoinformatics analysis and in silico designing of a novel multi-epitope peptide vaccine against Staphylococcus aureus. Infect Genet Evol. 2017;48: 83–94.

doi:10.1016/j.meegid.2016.12.010

14. Nezafat N, Karimi Z, Eslami M, Mohkam M, Zandian S, Ghasemi Y. Designing an efficient multi-epitope peptide vaccine against Vibrio cholerae via combined immunoinformatics and protein interaction based approaches. Comput Biol Chem. 2016;62: 82–95. doi:10.1016/j.compbiolchem.2016.04.006

15. Sanchez J, Holmgren J. Cholera toxin - a foe & a friend. Indian J Med Res. 2011;133: 153–163.

16. Taylor RK, Miller VL, Furlong DB, Mekalanos JJ. Use of phoA gene fusions to identify a pilus colonization factor coordinately regulated with cholera toxin. Proc Natl Acad Sci USA. 1987;84: 2833–2837. doi:10.1073/pnas.84.9.2833

17. Kirn TJ, Taylor RK. TcpF is a soluble colonization factor and protective antigen secreted by El Tor and classical O1 and O139 Vibrio cholerae serogroups. Infect Immun. 2005;73: 4461–4470. doi:10.1128/IAI.73.8.4461-4470.2005

18. Rashid MI, Rehman S, Ali A, Andleeb S. Fishing for vaccines against Vibrio cholerae using in silico pan-proteomic reverse vaccinology approach. PeerJ. 2019;7: e6223. doi:10.7717/peerj.6223

19. Sperandio V, Girón JA, Silveira WD, Kaper JB. The OmpU outer membrane protein, a potential adherence factor of Vibrio cholerae. Infect Immun. 1995;63: 4433–4438. doi:10.1128/iai.63.11.4433-4438.1995

20. Noureen N, Tariq M, Farooq A, Arif A, Bokhari H. In silico identification of receptor specific epitopes as potential vaccine candidates from Vibrio cholerae strains. Gene Rep. 2016;4: 222–232. doi:10.1016/j.genrep.2016.07.002

21. Zeb S, Ali A, Gulfam SM, Bokhari H. Preliminary Work Towards Finding Proteins as Potential Vaccine Candidates for Vibrio cholerae Pakistani Isolates through Reverse Vaccinology. Medicina (Kaunas). 2019;55. doi:10.3390/medicina55050195

22. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25: 3389–3402. doi:10.1093/nar/25.17.3389

23. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, et al. VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res. 2005;33: D325-8. doi:10.1093/nar/gki008

24. 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

25. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics. 2010;26: 1608–1615.

doi:10.1093/bioinformatics/btq249

26. Sengupta DK, Sengupta TK, Ghose AC. Major outer membrane proteins of Vibrio cholerae and their role in induction of protective immunity through inhibition of intestinal colonization. Infect Immun. 1992;60: 4848–4855. doi:10.1128/iai.60.11.4848-4855.1992

27. Pal P. Role of Cholera Toxin in *Vibrio cholerae* Infection in Humans - A Review. ILNS. 2014;22: 22–32. doi:10.18052/www.scipress.com/ILNS.22.22

28. Reidl J, Klose KE. *Vibrio cholerae* and cholera: out of the water and into the host. FEMS Microbiol Rev. 2002;26: 125–139. doi:10.1111/j.1574-6976.2002.tb00605.x

29. Nain Z, Abdulla F, Rahman MM, Karim MM, Khan MSA, Sayed SB, et al. Proteome-wide screening for designing a multi-epitope vaccine against emerging pathogen Elizabethkingia anophelis using immunoinformatic approaches. J Biomol Struct Dyn. 2020;38: 4850–4867. doi:10.1080/07391102.2019.1692072

30. Jalajakumari MB, Manning PA. Nucleotide sequence of the gene, ompW, encoding a 22kDa immunogenic outer membrane protein of Vibrio cholerae. Nucleic Acids Res. 1990;18: 2180. doi:10.1093/nar/18.8.2180

31. Fu X, Zhang J, Li T, Zhang M, Li J, Kan B. The Outer Membrane Protein OmpW Enhanced*V. cholerae*Growth in Hypersaline Conditions by Transporting Carnitine. Front Microbiol. 2017;8: 2703. doi:10.3389/fmicb.2017.02703

32. Bina JE, Mekalanos JJ. Vibrio cholerae tolC is required for bile resistance and colonization. Infect Immun. 2001;69: 4681–4685. doi:10.1128/IAI.69.7.4681-4685.2001

33. Xicohtencatl-Cortés J, Lyons S, Chaparro AP, Hernández DR, Saldaña Z, Ledesma MA, et al. Identification of proinflammatory flagellin proteins in supernatants of Vibrio cholerae O1 by proteomics analysis. Mol Cell Proteomics. 2006;5: 2374–2383. doi:10.1074/mcp.M600228-MCP200

34. Ikhtiar AM, Jazairi B, Khansa I, Othman A. HLA class I alleles frequencies in the Syrian population. BMC Res Notes. 2018;11: 324. doi:10.1186/s13104-018-3427-1

35. Jazairi B, Khansaa I, Ikhtiar A, Murad H. Frequency of HLA-DRB1 and HLA-DQB1 Alleles and Haplotype Association in Syrian Population. Immunol Invest. 2016;45: 172–179. doi:10.3109/08820139.2015.1131293

36. Human Leukocyte Antigen (HLA) System - Immunology; Allergic Disorders - MSD Manual Professional Edition. [cited 3 Nov 2022]. Available: https://www.msdmanuals.com/professional/immunology-allergic-disorders/biology-of-the-immune-system/human-leukocyte-antigen-hla-system

37. Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, et al. NetMHCpan, a method for MHC class I binding prediction beyond humans. Immunogenetics. 2009;61: 1–13. doi:10.1007/s00251-008-0341-z

38. Nielsen M, Lundegaard C, Blicher T, Lamberth K, Harndahl M, Justesen S, et al. NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. PLoS ONE. 2007;2: e796. doi:10.1371/journal.pone.0000796

39. Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. Nucleic Acids Res. 2020;48: W449–W454. doi:10.1093/nar/gkaa379

40. Zhang J, Tao A. Antigenicity, Immunogenicity, Allergenicity. In: Tao A, Raz E, editors. Allergy Bioinformatics. Dordrecht: Springer Netherlands; 2015. pp. 175–186. doi:10.1007/978-94-017-7444-4_11

41. Calis JJA, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A, et al. Properties of MHC class I presented peptides that enhance immunogenicity. PLoS Comput Biol. 2013;9: e1003266. doi:10.1371/journal.pcbi.1003266

42. Dimitrov I, Bangov I, Flower DR, Doytchinova I. AllerTOP v.2--a server for in silico prediction of allergens. J Mol Model. 2014;20: 2278. doi:10.1007/s00894-014-2278-5

43. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Raghava GPS. Peptide toxicity prediction. Methods Mol Biol. 2015;1268: 143–157. doi:10.1007/978-1-4939-2285-7_7

44. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Open Source Drug Discovery Consortium, et al. In silico approach for predicting toxicity of peptides and proteins. PLoS ONE. 2013;8: e73957. doi:10.1371/journal.pone.0073957

45. Jensen KK, Andreatta M, Marcatili P, Buus S, Greenbaum JA, Yan Z, et al. Improved methods for predicting peptide binding affinity to MHC class II molecules. Immunology. 2018;154: 394–406. doi:10.1111/imm.12889

46. Nielsen M, Lund O. NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. BMC Bioinformatics. 2009;10: 296. doi:10.1186/1471-2105-10-296

47. Chaplin DD. Overview of the immune response. J Allergy Clin Immunol. 2010;125: S3–S23. doi:10.1016/j.jaci.2009.12.980

48. Dhanda SK, Gupta S, Vir P, Raghava GPS. Prediction of IL4 inducing peptides. Clin Dev Immunol. 2013;2013: 263952. doi:10.1155/2013/263952

49. Saha S, Raghava GPS. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins. 2006;65: 40–48. doi:10.1002/prot.21078

50. Saha S, Raghava GPS. Prediction methods for B-cell epitopes. Methods Mol Biol. 2007;409: 387–394. doi:10.1007/978-1-60327-118-9_29

51. Bui H-H, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population

coverage of T-cell epitope-based diagnostics and vaccines. BMC Bioinformatics. 2006;7: 153. doi:10.1186/1471-2105-7-153

52. Bastola R, Noh G, Keum T, Bashyal S, Seo J-E, Choi J, et al. Vaccine adjuvants: smart components to boost the immune system. Arch Pharm Res. 2017;40: 1238–1248. doi:10.1007/s12272-017-0969-z

53. Ayyagari VS, T C V, K AP, Srirama K. Design of a multi-epitope-based vaccine targeting M-protein of SARS-CoV2: an immunoinformatics approach. J Biomol Struct Dyn. 2022;40: 2963–2977. doi:10.1080/07391102.2020.1850357

54. Li G, Huang Z, Zhang C, Dong B-J, Guo R-H, Yue H-W, et al. Construction of a linker library with widely controllable flexibility for fusion protein design. Appl Microbiol Biotechnol. 2016;100: 215–225. doi:10.1007/s00253-015-6985-3

55. He J, Huang F, Zhang J, Chen Q, Zheng Z, Zhou Q, et al. Vaccine design based on 16 epitopes of SARS-CoV-2 spike protein. J Med Virol. 2021;93: 2115–2131. doi:10.1002/jmv.26596

56. Klein JS, Jiang S, Galimidi RP, Keeffe JR, Bjorkman PJ. Design and characterization of structured protein linkers with differing flexibilities. Protein Eng Des Sel. 2014;27: 325–330. doi:10.1093/protein/gzu043

57. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, et al. Protein identification and analysis tools in the ExPASy server. Methods Mol Biol. 1999;112: 531–552. doi:10.1385/1-59259-584-7:531

58. Hebditch M, Carballo-Amador MA, Charonis S, Curtis R, Warwicker J. Protein-Sol: a web tool for predicting protein solubility from sequence. Bioinformatics. 2017;33: 3098–3100. doi:10.1093/bioinformatics/btx345

59. Magnan CN, Randall A, Baldi P. SOLpro: accurate sequence-based prediction of protein solubility. Bioinformatics. 2009;25: 2200–2207. doi:10.1093/bioinformatics/btp386

60. Paraskevopoulou V, Falcone FH. Polyionic tags as enhancers of protein solubility in recombinant protein expression. Microorganisms. 2018;6. doi:10.3390/microorganisms6020047

61. Han X, Ning W, Ma X, Wang X, Zhou K. Improving protein solubility and activity by introducing small peptide tags designed with machine learning models. Metab Eng Commun. 2020;11: e00138. doi:10.1016/j.mec.2020.e00138

62. Combet C, Blanchet C, Geourjon C, Deléage G. NPS@: network protein sequence analysis. Trends Biochem Sci. 2000;25: 147–150. doi:10.1016/s0968-0004(99)01540-6

63. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 2015;10: 845–858. doi:10.1038/nprot.2015.053

64. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr. 1993;26: 283–291. doi:10.1107/S0021889892009944

65. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 2007;35: W407-10. doi:10.1093/nar/gkm290

66. Sippl MJ. Recognition of errors in three-dimensional structures of proteins. Proteins. 1993;17: 355–362. doi:10.1002/prot.340170404

67. Rapin N, Lund O, Bernaschi M, Castiglione F. Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. PLoS ONE. 2010;5: e9862. doi:10.1371/journal.pone.0009862

68. Ponomarenko J, Bui H-H, 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

69. Biselli E, Schink SJ, Gerland U. Slower growth of Escherichia coli leads to longer survival in carbon starvation due to a decrease in the maintenance rate. Mol Syst Biol. 2020;16: e9478. doi:10.15252/msb.20209478

70. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic Acids Res. 2005;33: W363-7. doi:10.1093/nar/gki481

71. Duhovny D, Nussinov R, Wolfson HJ. Efficient unbound docking of rigid molecules. In: Guigó R, Gusfield D, editors. Algorithms in Bioinformatics. Berlin, Heidelberg: Springer Berlin Heidelberg; 2002. pp. 185–200. doi:10.1007/3-540-45784-4_14

72. Ahammad I, Lira SS. Designing a novel mRNA vaccine against SARS-CoV-2: An immunoinformatics approach. Int J Biol Macromol. 2020;162: 820–837. doi:10.1016/j.ijbiomac.2020.06.213

73. Raheem SG, Salh KK, Ibrahim KS, Gorji AE. In-Silico Designing a Multi-Peptide Vaccine: Against Vibrio Cholera. Int J Pept Res Ther. 2021;27: 1541–1553. doi:10.1007/s10989-021-10190-3

74. Enayatkhani M, Hasaniazad M, Faezi S, Gouklani H, Davoodian P, Ahmadi N, et al. Reverse vaccinology approach to design a novel multi-epitope vaccine candidate against COVID-19: an in silico study. J Biomol Struct Dyn. 2021;39: 2857–2872. doi:10.1080/07391102.2020.1756411

75. Ko J, Park H, Heo L, Seok C. GalaxyWEB server for protein structure prediction and refinement. Nucleic Acids Res. 2012;40: W294-7. doi:10.1093/nar/gks493

76. Connell TD. Cholera toxin, LT-I, LT-IIa and LT-IIb: the critical role of ganglioside binding

in immunomodulation by type I and type II heat-labile enterotoxins. Expert Rev Vaccines. 2007;6: 821–834. doi:10.1586/14760584.6.5.821

77. Andrusier N, Nussinov R, Wolfson HJ. FireDock: fast interaction refinement in molecular docking. Proteins. 2007;69: 139–159. doi:10.1002/prot.21495

78. Mashiach E, Schneidman-Duhovny D, Andrusier N, Nussinov R, Wolfson HJ. FireDock: a web server for fast interaction refinement in molecular docking. Nucleic Acids Res. 2008;36: W229-32. doi:10.1093/nar/gkn186

79. Pulendran B. Learning immunology from the yellow fever vaccine: innate immunity to systems vaccinology. Nat Rev Immunol. 2009;9: 741–747. doi:10.1038/nri2629

80. Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC, et al. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic Acids Res. 2005;33: W526-31. doi:10.1093/nar/gki376

81. Using one-step DNA assembly to create designer DNA. [cited 25 Dec 2022]. Available: https://www.nature.com/articles/d43747-020-00923-6