Original Research

Immunoinformatic and dynamic simulation-based designing of a multi-epitope vaccine against emerging pathogen *Elizabethkingia anophelis*

Zulkar Nain¹, Faruq Abdullah², M. Mizanur Rahman¹, Mohammad Minnatul Karim¹, S. M. Raihan Rahman¹, Shakil Ahmed Khan¹, Sifat Bin Sayed¹, Md. Moinuddin Sheam¹, Zahurul Haque¹, Utpal Kumar Adhikari^{3*}

¹Department of Biotechnology and Genetic Engineering, Faculty of Biological Sciences, Islamic University, Kushtia-7003, Bangladesh.

²Department of Statistics, Faculty of Sciences, Islamic University, Kushtia-7003, Bangladesh.

³School of Medicine, Western Sydney University, Campbelltown, NSW-2560, Australia.

*Corresponding author (u.adhikari@westernsydney.edu.au)

Abstract

Elizabethkingia anophelis is an emerging human pathogen causing neonatal meningitis, catheter-associated infections and nosocomial outbreaks with high mortality rates. Besides, they are resistant to most antibiotics used in empirical therapy. In this study, therefore, we used immunoinformatic approaches to design an epitope-based vaccine against *E. anophelis* as an alternative preventive measure. Initially, T-cell (CTL and HTL) and B-cell (LBL) epitopes were predicted from the highest antigenic protein. The CTL and HTL epitopes together had a population coverage of 99.97% around the world. Eventually, 6 CTL, 7 HTL, and 2 LBL epitopes were selected and used to construct a multiepitope vaccine. The vaccine protein was found to be highly immunogenic, non-allergenic, and non-toxic. Codon adaptation and *in silico* cloning were performed to ensure better expression within *E. coli* K12 host system. The stability of the vaccine structure was also improved by disulphide bridging. In addition, molecular docking and dynamic simulation revealed good and stable binding affinity between the vaccine and receptor. The immune simulation showed higher levels of T-cell and B-cell activities which was in coherence with actual immune response. Repeated exposure simulation resulted in higher clonal selection and faster antigen clearance. Nevertheless, experimental validation is required to ensure the immunogenic potency and safety of this vaccine to control *E. anophelis* infection in the future.

Keyword: *Elizabethkingia anopheles*; Immunoinformatics; Multiepitope Vaccine; Dynamic Simulation; Immune Simulation.

Background

Elizabethkingia anophelis is an opportunistic human pathogen responsible for neonatal meningitis^{1,2} and infections in immunocompromised^{3,4} and postoperative patients⁵. These are Gram-negative bacilli belonging to the family of *Flavobacteriaceae*^{1,6}. Despite their widespread presence¹ and rare association with human infection, it has recently been reported that bacteria of the genus *Elizabethkingia* are involved in life-threatening infections⁷. The first isolation of *E. anophelis* from the midgut of mosquito *Anopheles gambiae* was dated back to 2011⁸. Over the last five years, *E. anophelis* has caused multiple outbreaks in South Africa⁹, Singapore¹⁰, Hong Kong¹¹, Taiwan⁷ and the United States¹² with a mortality of up to 60%^{7,11,13}. For instance, two large outbreaks in the Midwestern United States resulted in a death toll of 31% (20/65) in 65 cases^{13,14}. In addition, in Illinois, at least 10 cases of *E. anophelis* infections were reported from January 2014 to April 2016 with six (60%) of the patients dying from this infection³. The significant mortality outcome from *E.*

anophelis infections may be correlated with the virulence factors (*i.e.*, capsule, lipopolysaccharide, magnesium transport protein, etc.) of the bacteria and pre-existing health conditions of the sufferers ⁷. Also, multiple antimicrobial resistance is an integral part of the *E. anophelis* genome, rendering it immune to the last line of antibiotics used in empirical therapy ¹.

Under the genus *Elizabethkingia*, *E. meningoseptica* is well-documented as a deadly infectious agent, which can also cause meningitis and sepsis as well as nosocomial outbreaks with a mortality rate of about 50% ⁵. Interestingly, recent studies indicated that many of the previously described *E. meningoseptica* were indeed *E. anophelis* ⁷. Therefore, a significant portion of the previous *Elizabethkingia* infections could actually be *E. anophelis* infections ¹¹. As a result, *E. anophelis* has been considered as an emerging opportunistic pathogen whose clinical significance has recently been emphasized ¹. However, little is known about this newly emerging pathogen. For instance, the transmission route of *E. anophelis* still remains elusive ^{6,15}. These altogether have made conventional treatment against *E. anophelis* infections obsolete and challenging ². Nevertheless, there are no vaccines or treatments available to prevent and/or control the diseases associated with *E. anophelis*. Therefore, alternative control measures are essential to counteract their pathogenic potential.

Peptide vaccines are a newer form of immune stimulant where fragments (*i.e.*, epitopes) of antigenic protein are used to mimic the natural presence of pathogens ¹⁶. In 1985, the first epitope-based vaccine was developed against cholera in *Escherichia coli* ¹⁷. Besides, there are many peptide vaccines under development, such as the vaccine for HIV, malaria, swine fever, influenza, anthrax, etc. ¹⁶. The developmental process includes the identification of the virulence protein, selection of peptide segments and generation of cellular and humoral immune response ¹⁸. The protein antigen in *E. anophelis* proteome can be identified from the recently sequenced *E. anophelis* Ag1 genome ¹⁹. Interestingly, computer-based advanced approaches like immunoinformatics can be useful in this scenario. In recent years, the use of immunoinformatics to design *in silico* peptide vaccine model against viruses ^{20–26} bacteria ^{27–33} and parasites ^{34–36} have been intensified. Therefore, peptide vaccine can be a more attractive alternative as preventive means against *E. anophelis*.

The current study involves the screening of the *E. anophelis* proteome to identify the highest antigenic protein, followed by the prediction of different T and B-cell epitopes with their corresponding MHC alleles. These epitopes were further evaluated prior to estimating population coverage. Therefore, a multiepitope vaccine was designed using the most potent epitopes with appropriate adjuvant and linkers. The primary sequence was used for the immunogenic and physiochemical profiling of the vaccine protein and for the prediction of secondary and tertiary structure. The predicted 3D structure was then refined and subjected to disulphide engineering to improve stability. The binding interaction of the vaccine to the receptor was calculated by molecular docking. Finally, molecular dynamics and immune simulation were performed to estimate the stability of the vaccine and its immunogenic potency in real-life, respectively. The overall computational workflow of the study is graphically illustrated in Fig. 1.

Prediction and assessment of CTL, HTL and LBL epitopes. A total of 105 unique CTL epitopes (9-mer) were predicted from the *E. anophelis* TSPO protein using the NetCTL 1.2 server. Among them, only 49 epitopes were found as antigenic, immunogenic, non-allergenic and non-toxic as tabulated in Supplementary Table 1. These epitopes were further used to predict their MHC-I binding alleles using MHC-I allele prediction tool of the IEDB server. Similarly, a total of 136 unique HTL epitopes (15-mer) and their MHC-II binding molecules were predicted using the IEDB MHC-II prediction tool. Further, the cytokine (*i.e.*, IFN- γ , IL-4, and IL-10) inducing ability of these HTL epitopes were evaluated. However, only 17 HTL epitopes were able to induce all three types of cytokines as shown in Supplementary Table 2. B-cell epitopes are antigenic regions of a protein that can trigger antibody formation. We found 4 linear B lymphocyte (LBL) epitopes in of variable length ranged in 8 to 26 amino acid residues while their scores ranged from 0.686 to 0.794. All predicted LBL epitopes, along with their length and scores, are provided in Supplementary Table 3. After evaluation, however, only 2 LBL epitopes (*i.e.*, MKRQIDWKALAA and ALTPANSFNWYNQLDKPLFTPPSNF) were found to be antigenic, non-allergenic and non-toxic, hence, considered for vaccine construction.

Table 2: Final selected helper T lymphocyte (HTL) epitopes with their cytokine inducing properties.

HTL Epitopes	Position	Common Core	Interferon- γ	Interleukin-4	Interleukin-10
AFYFIWRAKENRFBKH	66-80	WRAKENRFBK	Positive	IL4 inducer	IL10 inducer
FIWRAKENRFBKHRQK	69-83	WRAKENRFBK	Positive	IL4 inducer	IL10 inducer
FYFIWRAKENRFBKHR	67-81	WRAKENRFBK	Positive	IL4 inducer	IL10 inducer
IAFYFIWRAKENRFBK	65-79	WRAKENRFBK	Positive	IL4 inducer	IL10 inducer
IWRAKENRFBKHRQKA	70-84	WRAKENRFBK	Positive	IL4 inducer	IL10 inducer
YFIWRAKENRFBKHRQ	68-82	WRAKENRFBK	Positive	IL4 inducer	IL10 inducer
WRAKENRFBKHRQKAL	71-85	WRAKENRFBK	Positive	IL4 inducer	IL10 inducer

Population coverage analysis. The finally selected 49 CTL and 17 HTL epitopes and their corresponding HLA alleles were used to estimate the population coverage both individually and in combination as tabulated in Supplementary Table 4. Our selected CTL and HTL epitopes were found to cover 99.92% and 68.89% of the world population, respectively. In this study, we attached importance to the combined population coverage, as both CTL and HTL epitopes were considered in the vaccine to construct. Importantly, CTL and HTL epitopes showed 99.97% population coverage when used in combination. The highest population coverage (100%) was found for several European countries (*viz.*, Bulgaria, England, Finland, France, Germany, Poland, and Sweden) and the South American country Peru. The least population coverage (0%) was predicted for the European country Slovakia, which is plausible due to its lower allele frequency (based on the data available in Allele Frequencies (<http://allelefrequencies.net>) database). In West Africa, where the bacteria first appeared, the population coverage was 99.81%. *E. anophelis* caused several outbreaks in different countries of the world, especially in the USA. Therefore, population coverage in these regions is important for vaccine development. Importantly, the cumulative population coverages found in the outbreak region USA, South Africa, China, Singapore, Hong Kong, and Taiwan were 99.98%, 99.51%, 99.39%, 99.44%, 97.96%, and 99.71%, respectively as depicted in Fig. 2. The result suggested that our vaccine could help to control *E. anophelis* in these regions. In addition, the population coverage in South Asia and Australia were found to be 99.86% and 99.30%, respectively.

Figure 2: Worldwide population coverage by selected T-cell epitopes and their respective MHC binding alleles. Regions of particular importance were considered in this graph where pink and yellow circles are indicating the regions of outbreaks and bacterial origin, respectively.

Molecular docking between epitopes and HLA alleles. In general, the epitopes having affinity to multiple alleles are better candidates for vaccine design because of high population coverage. Therefore, we considered 6 CTL and 7 HTL epitopes which have interactions to ≥ 5 of their respective alleles with an experimentally validated allele in common. We found that 6 CTL epitopes shared a common allele HLA-B*53:01, while 7 HTL epitopes shared an HLA-DRB5*01:01 allele as shown in Supplementary Table 5. So, these 13 epitopes (6 CTLs and 7 HTLs) were modelled using PEPFOLD 3 server. Finally, the molecular docking between these epitopic models and their respective common alleles was performed with AutoDock tools and AutoDock Vina. The binding affinities between epitopes and their binding alleles were compared to the control ligand which is given in Supplementary Table 6 and Figure 1. The best interacting CTL epitope model was LISLWFIIF (-10.0 kcal/mol) as compared to the CTL positive control (-9.3 kcal/mol) and the residues involved in classic hydrogen bond interactions were Leu1, Ile2, Ile7, Phe9, Asn63, Asn70, Asn77, Arg97, Tyr99, Thr143, Tyr123, Tyr9, Tyr171, Trp5, and Ser3 as shown in Fig. 3A. On the other hand, the HTL epitope model YFIWRAKENRFRKHRQ was the best (-7.3 kcal/mol) as compared to HTL positive control (-8.5 kcal/mol). The conventional hydrogen bond network in YFIWRAKENRFRKHRQ comprised of residue Gln9, Ser53, Glu55, Asn62, Tyr213, His281, Asn282, Arg5, Asn9, Arg14, Gln15, Glu8, His13, Phe2, Ala6, Phe51, and Asp270 which are shown in Fig. 3B. Thus, these 13 T-cell epitopes ensured their binding efficiency as well as their suitability to be used in multi-epitope based vaccine design.

Figure 3: Molecular docking between T cell epitopes and their respective MHC alleles. (A) Molecular docking between best CTL epitope (LISLWFIIF) and HLA-B*53:01 allele where residue Leu1, Ile2, Ile7, Phe9, Asn63, Asn70, Asn77, Arg97, Tyr99, Thr143, Tyr123, Tyr9, Tyr171, Trp5, and Ser3 were involved in conventional hydrogen bonds, (B) Molecular docking between best HTL epitope (YFIWRAKENRFKHRQ) and HLA-DRB5*01:01 where residue Gln9, Ser53, Glu55, Asn62, Tyr213, His281, Asn282, Arg5, Asn9, Arg14, Gln15, Glu8, His13, Phe2, Ala6, Phe51, and Asp270 took part in conventional hydrogen bonds.

Construction of multi-epitope vaccine. In multi-epitope vaccine design, we have considered highly potent T-cell epitopes that share at least one of their respective allele and B-cell epitopes that are antigenic, non-allergenic and non-toxic. Thus, a multi-epitope vaccine is constructed using 7 HTL, 6 CTL and 2 BL epitopes fused by GPGPG, AAY, and KK linkers, respectively (Table 1 and 2). Epitopes in each group were sequenced according to their higher to lower antigenic score. Furthermore, TLR4 agonist 50S ribosomal protein L7/L12 (130 residues) was added as an adjuvant using EAAAK linkers. The arrangement of different epitopes along with their joining linkers is shown in Fig. 4. The final vaccine construct comprises 384 amino acid residues.

Figure 4: Graphical illustration of multi-epitope vaccine construct. The final vaccine construct is 384 amino acid residues long in which first 130 residues have been represented as adjuvant followed by HTL (131–269 amino acids), CTL (270–342 amino acids) and BCL (343–384) epitopes. Adjuvant and HTL epitope has been joined by EAAAK linker (maroon) while GPGPG (blue), AAY (yellow) and KK (light green) linkers were used to join the CTL, HTL and BL epitopes, respectively.

Evaluation of the primary sequence of vaccine protein. The physiochemical and immunogenic nature of chimera protein was evaluated as shown in Table 3. The final vaccine construct was found to be basic, stable, thermostable, hydrophilic and highly soluble based on the analysis of physiochemical parameters. In addition, the half-life was estimated to be 30 hours in mammalian reticulocytes (*in vitro*), over 20 hours in yeast (*in vivo*) and over 10 hours in *E. coli* (*in vivo*). The immunogenic assessment also revealed that our vaccine construct is highly antigenic, immunogenic and non-allergenic (Table 3). These results indicated that this chimera protein is suitable to be a potential vaccine candidate.

Table 3: Antigenic, allergenic and physiochemical assessments of the primary sequence of final vaccine protein.

Evaluating Parameters	Results	Remarks
Number of amino acids	384	Suitable
Molecular weight	43754.20 Dalton	Average
Chemical formula	C ₂₀₇₉ H ₃₁₃₃ N ₅₃₃ O ₅₀₄ S ₃	-
Theoretical pI	9.99	Basic
Extinction coefficient (at 280 nm in H ₂ O)	87890 M ⁻¹ cm ⁻¹	-
Estimated half-life (mammalian reticulocytes, <i>in vitro</i>)	30 hours	-
Estimated half-life (yeast cells, <i>in vivo</i>)	>20 hours	-
Estimated half-life (<i>Escherichia coli</i> , <i>in vivo</i>)	>30 hours	-
Instability index	17.95	Stable
Aliphatic index	83.31	Thermostable
Grand average of hydropathicity (GRAVY)	-0.076	Hydrophilic
Antigenicity	0.9258	Antigenic
Immunogenicity	7.034	Immunogenic
Allergenicity	Probable non-allergen	Non-allergenic
Solubility	0.72056	Soluble

Secondary and tertiary structure prediction. The secondary structure was predicted using the latest PSIPRED 4.0 server which demonstrated the presence of 54.17% alpha helix (208), 8.59% beta strand (33) and 37.24% random coil (143) structure in our vaccine protein as depicted in Fig. 5A. Furthermore, homology modelling of vaccine protein was performed using the RaptorX server. A total of 384 amino acid residues was modelled as three domain(s) with 14% disorder (Fig. 5B). The RaptorX server uses multi-template based approach to build a protein model. For example, the template PDB ID: 1dd3A was chosen to model domain 1 (1-135 AA). Similarly, for domain 2 (232-384 AA) PDB ID: 5ujaA, 5u71A and 5tsiA; and domain 3 (136-231 AA) PDB ID: 5vflL were considered. P-value is a quality indicator in homology modelling and low P-value defines the high-quality model. The P-value obtained for the modelled structure was 8.67e-04, which is low and significant (Supplementary Table 7).

Figure 5: Prediction and assessment of the secondary and tertiary structure of the vaccine construct. (A) cartooned secondary structure of final vaccine construct showing α -helix (54.17%), β -strands (8.59%) and coils (37.24%), (B) 3D structure of vaccine protein showing three domains in different colours, (C) ProSA validation of 3D vaccine model showing Z-score (-2.54), and (d) Ramachandran plot analysis of refined structure showing 94.2%, 5.2% and 0.5% residues in favoured, allowed and disallowed region, respectively.

Refinement and validation of tertiary structure. Structure refinement with GalaxyRefine leads to an increase in the number of residues in the favoured region. Of all the refined models, the model 3 proved to be the best on the basis of various parameters including GDT-HA (0.9206), RMSD (0.527), MolProbity (2.502), Clash score (23.9), Poor rotamers (2.0) and Rama favoured (94.0) as shown in Supplementary Table 8. This model was taken as the final vaccine model and subjected to structural validation by plotting a Ramachandran plot and found the same that 94.2% (360) residue in Rama-favoured region, 5.2% (20) residues in allowed region and only 0.5% (2) residues in outlier region (Fig. 5D). ProSA-web verified the quality and potential errors in a crude 3D model. The overall quality factor of a modelled protein was 82.2% using ERRAT. While ProSA-web has shown the Z-score of -2.54 for the input vaccine protein model which is lying outside the scores range that commonly found in the case of native proteins of comparable size (Fig. 5C).

Mapping of B-cell epitopes in vaccine construct. The presence of B-cell epitopes was predicted using the ElliPro tool of IEBD, allowing the identification of 9 linear B-cell epitopes with a score of 0.527 to 0.819 and 10 conformational B-cell epitopes comprising of a total of 205 amino acid residues as shown in Fig. 6A and Fig. 6B, respectively. The scores of the conformational B-cell epitopes ranged from 0.536 to 0.920, while the size ranged from 3 to 64 residues.

Figure 6: B lymphocyte epitope present in the vaccine: (A) Red spheres showing linear BL epitopes containing 1. 24 residues with 0.819, 2. 5 residues with 0.803, 3. 16 residues with 0.655, 4. 14 residues with 0.654, 5. 19 residues with 0.552, 6. 11 residues with 0.543, 7. 85 residues with 0.805, 8. 8 residues with 0.628 and 9. 4 residues with 0.527 score); and (B) Green spheres showing conformational BL epitopes containing 1. 3 residues with 0.92, 2. 30 residues with 0.833, 3. 21 residues with 0.805, 4. 6 residues with 0.78, 5. 64 residues with 0.743, 6. 10 residues with 0.671, 7. 7 residues with 0.633, 8. 59 residues with 0.59, 9. 7 residues with 0.56, and 10. 3 residues with 0.532 score).

Disulphide bridging for vaccine stability. A total of 35 pairs of residues were found that could be used for disulphide engineering which is given in Supplementary Table 9. However, after evaluating other parameters such as energy and χ^3 , only a pair of residuals were finalized because their value falls below the allowable range *i.e.*, the value of energy should be less than 2.2 kcal/mol and χ^3 angle should be in between -87 and $+97$ degree³⁷. Therefore, a total of two mutations were generated on the residue pairs, Gly173-Glu203, for which the χ^3 angle and the energy were -86.78 degree and 1.91 kcal/mol, respectively as pointed in Fig. 7.

Figure 7: Disulphide engineering of the final vaccine protein. Mutated residue pairs showed in pink (initial residue: Gly173) and blue colour (initial residue: Glu203). Residues were selected based on their energy, χ^3 value, and B-factor.

Codon adaptation and *in silico* cloning. The main purpose of *in silico* cloning was to express the *E. anophelis*-derived vaccine protein epitope into the *E. coli* expression system. Therefore, it was necessary to adapt the codon respective to the subunit vaccine construct as per the codon usage of *E. coli* expression system. We adapted the codons according to the *E. coli* K12 strain using JCAT server and found that the GC-content of the improved sequence was 49.39% while the value of the codon adaptive index was 0.95 close to 1.0. Therefore, a satisfactory adaptation was achieved. Later on, XhoI and NdeI restriction sites were created and cloned into the vector pET28a (+) as shown in Fig. 8. The target sequence is also included at both ends between 6-histidine tag residues, which is helpful for purification purposes. Thus, the total length of the clone was 6.45 kbp.

Figure 8: *In silico* cloning. Vaccine protein sequence adapted to pET28a (+) vector showing the region of choice in blue colour surrounded between XhoI (158) and NdeI (1317) while the vector has shown in black lines.

Molecular docking of vaccine with TLR4 receptor. A total of 30 models of the vaccine-receptor complex was generated in molecular docking. Energy scores for these models are tabulated in Supplementary Table 10. Among these occupied models, the receptor was appropriately located and had the lowest energy value for model number 5, which was therefore selected as the best-docked complex (Fig. 9A). The energy score obtained for model 5 was found to be -1626.2 which is lowest among all predicted docked complexes and thus, shows the highest binding interaction.

Dynamic simulation of the vaccine-TLR4 complex. The vaccine-TLR4 complex becomes stabilized around 4 ns in the 5 ns long dynamic simulation as shown in Fig. 9B. The average simulation energy was -7261872.45 kJ/mol, while the Coulombic charge and van der Waals interactions were -9968104.77 and 1284708.44 kJ/mol, respectively (Supplementary Table 11). The RMSD was calculated to determine the structural stability of the vaccine-TLR4 complex. The average RMSD of the complex backbone was 3.55 Å while the RMSD value of side-chain residue varied between 0.73 Å to 24.44 Å (Fig. 9C). Furthermore, RMSF was observed to evaluate the side-chain flexibility. The RMSF value ranged from 0.61 Å to 10.68 Å and the higher peaks at ~10.68 Å in the plot indicate highly flexible regions in the complex (Fig. 9C). The illustration of the initial and final state of the complex demonstrated that both vaccine and TLR4 receptor have remarkable changes in their

conformation to facilitate the binding to form a stable complex in the thermodynamic environment (Fig. 9D). In addition, the number of initial and final hydrogen bonds were 1019 and 675, respectively. On average, 650 hydrogen bonds were present during the simulation, suggesting strong binding of the vaccine-receptor complex. These values indicate strong complex stability based on previous observations^{34,35}. Slight fluctuations in amino acid side chains also indicate the uninterrupted interaction between vaccine and receptor.

Figure 9: Molecular docking and dynamic simulation studies. (A) Molecular docking between vaccine (sphere) and TLR4 receptor, (B) RMSD value of the vaccine-TLR4 complex backbone; (C) Plot showing the RMSF (red) and RMSD (black) values of side-chain residues, and (D) Conformational illustration of initial (blue) and final (pink) step of vaccine-TLR4 complex in 5 ns long dynamic simulation.

***In silico* immune response simulation.** The *in silico* immune response generated in C-ImmSim immune simulator was consistent with actual immune responses as depicted in Fig. 10. The secondary and tertiary responses were significantly higher than the primary response. The normal high levels of immunoglobulin activity (*i.e.*, IgG1 + IgG2, IgM, and IgG + IgM antibodies) was evident in both secondary and tertiary responses with a corresponding decrease in antigen concentration. In addition, several B-cell isotypes with long-lasting activity were observed, indicating possible isotype switching potentials and memory formation (Fig. 10A-C). A similarly high response was seen in the TH (helper) and TC (cytotoxic) cell populations with corresponding memory development (Fig. 10D-F) which is essential for complementing the immune response. During exposure, increased macrophage activity was demonstrated while dendritic cell activity was found to be consistent (Fig. 10G-H). High levels of IFN- γ and IL-2 were also evident. In addition, a lower Simpson index (D) indicates greater diversity

(Fig. 10I). Moreover, repeated exposure to 12 injections elicited increasing IgG1 and decreasing IgM levels, while maintaining IFN- γ and TH cell populations at a high level throughout exposure as provided in Supplementary Figure 2. This profile indicates the development of immune memory and, consequently the increased clearance of the antigen at subsequent exposures.

Figure 10: *In silico* simulation of immune response using vaccine as antigen: (A) Antigen and immunoglobulins, (B) B cell population, (C) B cell population per state, (D) TH cell population, (E) TH cell population per state, (F) TC cell population per state, (G) MA population per state, (H) DC population per state, and (I) Production of cytokine and interleukins with Simpson index (D).

Discussion

Currently, vaccines are the most promising mode that improves the immune system²⁷. However, the development and production of the vaccine are labour-intensive and costly³⁸. Immunoinformatics can reduce this burden where a variety of tools and servers are applied to identify potential B and T-cell epitopes, which subsequently resulted in the designing of epitope-based subunit vaccine. Despite the recent outbreaks worldwide, there is no preventive measure that can suppress *E. anophelis*. In this study, the translocator protein (TSPO) in the *E. anophelis* proteome showed the highest antigenicity. TSPO belongs to a membrane protein family (Pfam: TspO/MBR family PF03073) and is associated with stress, membrane transport, and some neurodegenerative diseases³⁹. Therefore, the higher antigenic nature of this protein is no wonder in this case. Furthermore, we found that *E. anophelis*-

derived TSPO protein is not homologous to its human counterpart or any other proteins of human origin. Therefore, TSPO is a perfect target for the design and testing of a multi-epitope vaccine against *E. anophelis*.

The final vaccine protein comprised of antigenic, non-allergenic, and non-toxic B-cell and T-cell epitopes with higher binding affinity for several MHC molecules, which in turn provided high population coverage (99.99%) worldwide. It is important that it covers more than 97.5% population of *E. anophelis* outbreak regions. Besides, the highly antigenic and non-allergenic property of the final vaccine protein make it a potent vaccine construct. The molecular weight of vaccine protein was 54.62 kDa which is an average molecular weight for a multisubunit vaccine. Further, the theoretical pI indicates that our vaccine protein is basic and is most stable in the pH range of 9.04. Furthermore, the aliphatic index and instability index classify our vaccine as thermostable, while the negative GRAVY value suggests that the vaccine protein is hydrophilic and has strong interactions with water molecules. However, one of the biggest drawbacks in the field of therapeutic proteins is the short half-life of peptides⁴⁰. Interestingly, our vaccine protein half-life in mammalian reticulocytes (*in vitro*) was 30 hours and more than 20 hours and 10 hours in yeast and *E. coli* cells *in vivo*, respectively which is satisfactory based on earlier findings^{34,35}. Also, the recombinant proteins should be soluble on overexpression for post-production studies⁴¹. The predicted solubility of the vaccine protein ensured their ease access.

The tertiary structure of the vaccine was predicted and refined, followed by validation. The refinement of the crude model was ensured by their parametric scores. For instance, the crude model had 92.4% residues in Ramo favoured zone while refined models have up to 94.8% (Supplementary Table 8). In addition, GDT-HA score between 0.9 and 1.0 suggested great similarity of the refined models with the amino acid sequence. The Z-score (-2.54) and ERRAT quality factor (82.196) indicate that the overall quality of the refined vaccine structure was satisfactory. Besides, the Ramachandran plot shows that 94.2% of residues clustered tightly in the most favoured region with very few residues in outliers. A good quality model would probably exceed 90% in the most favoured regions⁴². Hence, the 3D structure of the vaccine is considered good. The expression of our recombinant vaccine protein inside *E. coli* K12 was improved through codon adaptation. The codon adaptation index (CAI) and GC content of the optimized DNA sequence were 0.95 and 49%, respectively. The CAI value (0.95) closer to 1.0 suggests the higher expression probability. The vaccine protein should be mechanically stable which is accomplished by creating unique disulphide bonds into multisubunit vaccine construct.

Further, molecular docking and dynamic simulation were performed to ensure the binding efficiency and stability of the vaccine-receptor complex, respectively. It is clear that the vaccine protein could properly occupy the TLR4 receptor with minimal energy. The simulated microscale-change in protein backbone (RMSD: 3.55 Å) and mild fluctuations in side-chain residues confirmed the stability of the vaccine-TLR4 complex. In addition, binding interactions within the complex were good according to hydrogen bond network analysis. Finally, *in silico* immune response simulation showed higher B and T-cell activity (helper T-cells in particular), and the results were in coherence with typical immune responses. The formation of B and T-cell memory was evident and last for several months. The high level of IFN- γ and IL-2 production during repeated exposure confirms the efficient Ig production, thereby, supporting a humoral response. Importantly, macrophages and dendritic cells are important when the expression of TLR4 in human is considered⁴³. In our study, both macrophage and dendritic cell activities were satisfactory. Moreover, components of the innate immune system such as epithelial cells were also active. The increase in Simpson index (D) over time indicates the emergence

of a dominating clone that means the higher the value, the lower the diversity⁴⁴. Therefore, the lower D value suggests a diverse immune response that is plausible considering the generated chimeric peptide which is composed of several B and T-cell epitopes.

However, there are certain limitations in immunoinformatics approach as well. For example, lack of standard benchmark, inadequate prediction methods, and unavailability of precise datasets for different computational analysis³². In addition, we have no idea about the exact receptor and path used by *E. anophelis* for their pathogenic entry, along with their specific immune-bypass mechanism within the human host. With this knowledge, the vaccine could be developed more effectively and efficiently to control and/or prevent the pathogenesis of *E. anophelis*.

Conclusion

E. anophelis has emerged as an opportunistic human pathogen involved in neonatal meningitis, sepsis, and several major nosocomial outbreaks. Despite their significant causality rate and antimicrobial resistance profile, the transmission route and control measure of this pathogen remain unattainable. In this study, we have utilized immunoinformatic approaches along with molecular dynamics and immune simulations to generate an effective multiepitope vaccine candidate. Therefore, the highest antigenic protein from *E. anophelis* proteome was selected followed by the prediction of T-cell and B-cell epitopes to produce cell-mediated and humoral immunity, respectively. Predicted epitopes were merged using suitable linkers and adjuvant to enhance the immunogenicity and adequate separation of epitopes within the human body. Allergenicity and antigenicity were also confirmed followed by the physicochemical properties evaluation. Molecular docking and dynamics were also performed to check the binding affinity and stability of TLR-4 and vaccine complex. Finally, disulphide engineering and *in silico* cloning were introduced to ensure the stability and effective expression of the vaccine construct. However, the proposed vaccine needs to be experimentally validated to ensure the control of *E. anophelis* infection by producing an active immunological memory.

Methodology

Retrieval of protein sequences and prediction of the highest antigenic protein. The FASTA formatted whole proteome of the *Elizabethkingia anophelis* Ag1, except protein homologs in *Homo sapiens*, was retrieved from the Integrated Microbial Genomes & Microbiomes 5.0 (IMG/M) system⁴⁵. It is heavily resourced by NCBI database⁴⁶ and integrates publicly available draft and complete genomes from all three domains of life⁴⁷. The 'Phylogenetic Profiler' tool of IMG/M system allows the identification of protein sequence for each gene in a query genome with or without the protein homologs in other genomes⁴⁷. Antigenicity is the ability to induce an immune response towards an antigen. Therefore, the selection of proteins with higher antigenicity is a prerequisite for peptide-based vaccine design. So, all proteins from *E. anophelis* proteome were evaluated using the VaxiJen 2.0 server with 0.5 threshold. This server maintains 70-89% prediction accuracy through auto cross-covariance (ACC) transformation⁴⁸.

Prediction of cytotoxic T lymphocyte (CTL) epitopes and their MHC-I binding alleles. Most cytotoxic T-cells express T-cell receptors (TCRs) that can recognize a specific antigen⁴⁹. Therefore, prediction of CTL epitopes is essential for coherent vaccine design. The selected protein was submitted to the NetCTL 1.2 server for the prediction of CTL epitopes with 0.5 threshold⁵⁰. This server predicts CTL epitopes based on three criteria, namely, MHC-I binding peptides, C terminal cleavage, and TAP transport efficiency. MHC-I binding and C-terminal cleavage are obtained by using artificial neural networks whereas TAP transporter efficiency is evaluated by the weight matrix

⁵⁰. Furthermore, MHC-I binding alleles for each 9-mer CTL epitope were predicted using the consensus method in the IEDB analysis tool ⁵¹. The MHC source species was human and percentile rank ≤ 2 was considered.

Assessment of CTL epitopes for immunogenicity, allergenicity, and toxicity. The suitability of CTL epitopes lies within their immunogenic potential, hence, considered as the centre of the vaccine adeptness ²⁶. So, all predicted CTL epitopes were evaluated for their immunogenicity using MHC-I immunogenicity tool of IEDB server ⁵². Further, their antigenicity is evaluated again to ensure their ability to induce immune response through VaxiJen 2.0 server ⁴⁸. Vaccine components should be free from allergic reaction. So, AllerTOP 2.0 server was used for allergenicity prediction. This server is designed to predict allergenicity based on κ -nearest neighbours (kNN) method with 88.7% accuracy ⁵³. Furthermore, epitopes having toxicity should be eliminated since they could compromise the aim and development of the vaccine. Therefore, we used ToxinPred server to identify the toxic CTL epitopes. This server works based on the different properties of peptides using machine learning technique and quantitative matrix ⁵⁴.

Prediction of helper T lymphocyte (HTL) epitopes and their MHC-II binding alleles. Helper T-lymphocytes (HTLs) are a part of the adaptive immune response and can generate cellular and humoral immune response against pathogens or foreign antigens. Therefore, the prediction of HTL epitopes which bind to MHC-II alleles are essential and play a significant role in rational vaccine design ⁵⁵. The HTL epitopes of 15-mer length (peptide) and their corresponding MHC class II alleles were predicted through the IEDB MHC II binding tool using the consensus method ⁵⁶. For the consensus method, a percentile rank is generated by comparing the peptide's IC50 against those of a set of random peptides from the SWISS-PROT database ⁵⁷. A lower percentile rank indicates high affinity, therefore, percentile rank ≤ 2 was considered in this study.

Identification of cytokine-inducing HTL epitopes. The helper T-cells help in the activation of cytotoxic T-cells and other immune cells upon releasing different types of cytokines *i.e.*, interferon-gamma (IFN- γ), interleukin-4 (IL-4) and interleukin-10 (IL-10) ⁵⁸. As a result, cytokine-inducing HTL epitopes are crucial for the development of vaccine or immunotherapy. Therefore, we used IFNepitope server to predict IFN- γ producing HTL epitopes with Motif and SVM based hybrid method and IFN- γ versus Non-IFN- γ model as the default parameters ⁵⁹. On the other hand, IL-4 and IL-10 inducing properties were predicted using IL4pred ⁶⁰ and IL10pred ⁶¹ servers, respectively. These two servers are designed for *in silico* prediction of interleukin inducing peptides. The IL4pred and IL10pred operations were carried out based on SVM method and hybrid method (Motif and SVM), respectively with other parameters kept as default.

Prediction and assessment of linear B lymphocyte (LBL) epitopes. B-cells provide humoral immunity by secreting immunoglobulins which can neutralize antigen upon binding ⁶². Hence, B-cell epitopes are also essential for effective vaccine construct. B-cell epitopes are two types, linear (continuous) and conformational (discontinuous). The ElliPro tool from the IEDB database can predict both types of B-cell epitopes based on flexibility and solvent-accessibility ⁶³. In case of vaccine design, however, only the linear epitopes are considered. Therefore, only LBL epitopes were predicted using the ElliPro tool with default parameters. The tool utilizes three different algorithms (*i.e.*, an approximation of protein shape, protrusion index (PI) of residues and clustering of neighbouring residues) where each predicted epitope was given a PI score. Residues with larger PI values are related to greater solvent accessibility ⁶³. The predicted LBL epitopes were subjected to

further evaluation in terms of their antigenic, allergenic and toxic nature using VaxiJen 2.0, AllerTOP 2.0 and ToxinPred server, respectively.

Estimation of population coverage. The distribution of HLA alleles, as well as their expression, could vary throughout the world according to the difference in regions and ethnicities²⁵. Therefore, successful vaccine development demands the evaluation of HLA allele distribution across the world population⁶⁴. In this study, the distribution of HLA alleles for potential CTL and HTL epitopes was evaluated with the IEDB population coverage tool⁶⁴. This server is designed to estimate the population coverage of epitopes from different regions based on the distribution of their MHC binding alleles. Since the prevalence of *E. anophelis* is poorly understood, the percentage of population coverage was calculated for different parts of the world. In particular, however, we emphasized the areas where it was first appeared and caused outbreaks.

Peptide modelling and molecular docking studies. Epitopes having a common corresponding allele with good binding interactions to more alleles are good choice for multiepitope vaccine design. Therefore, epitopes showing binding interaction to an experimentally validated common allele were targeted for molecular docking studies using the AutoDock tools⁶⁵ and AutoDock Vina software⁶⁶. Molecular docking is a way to measure the binding affinity between protein and ligand molecule. At first, the selected CTL and HTL epitopes were modeled using PEP-FOLD 3.0 server with sOPEP sorting scheme in 200 simulations⁶⁷. This server was designed to predict the conformations of small peptides (5-50 amino acids) based on Forward Backtrack/Taboo Sampling algorithm⁶⁷. For CTL epitopes, the most conserved allele HLA- B*53:01 protein (PDB ID: 1A1O) was downloaded from RCSB Protein Data Bank (PDB). The allele was then processed with Discovery Studio and AutoDock tools. The grid box centre was set at 5.78, 28.96 and 21.71 Å in the X, Y and Z axis, respectively while the size was set at 54, 42 and 28 Å in the X, Y and Z dimensions, respectively. Similarly, the allele HLA- DRB5*01:01 (PDB ID: 1HQR) was considered for HTL epitopes where grid box centre was set at 18.72, 27.28 and 28.80 Å in the X, Y and Z axis, respectively, and the size was set at 44, 38, and 34 Å in the X, Y and Z dimensions, respectively. The spacing was 1.0 Å for both cases. Finally, the docking simulation was performed using AutoDock Vina. The PyMOL molecular graphics system (PyMOL) was used to make docked complex and visualized in BIOVIA Discovery Studio 2017.

Designing of multiepitope vaccine construct. The multiepitope vaccine construct was designed by combining adjuvant, CTL, HTL and LBL epitopes with appropriate linkers as reported earlier^{27,36}. An adjuvant is an immunogenic composition that may boost the specific and long-lasting immune response against an antigen by a variety of mechanisms⁶⁸. Toll-like receptor (TLR) agonists are well-known for recognizing ligands from many Gram-negative bacteria and also to be a part of the peptide-based subunit vaccine³⁵. In addition, TLR4 receptor upregulates TSPO/MBR related proteins by overproducing cytokines⁶⁹. Therefore, a TLR4 agonist namely 50S ribosomal protein L7/L12 was retrieved from the NCBI database (Accession no. P9WHE3) and used as an adjuvant. For the effective separation of each epitope within the human body, EAAAK, GP GPG, AAY, and KK linkers were used. EAAAK linker was designed for effective separation of bifunctional fusion protein domains⁷⁰, therefore, utilized to join the adjuvant and first HTL epitope. For efficient recognition of multiepitope vaccine, GP GPG and AAY linkers were used to merge HTL and CTL epitopes. The bi-lysine (KK) linker was used to fuse B-cell epitopes to preserve their independent immunogenic activities⁷¹.

Immunogenic, allergenic and physiochemical evaluation of vaccine construct. Immunogenicity is the ability to stimulate cellular and humoral immune response whereas antigenicity is the ability to

recognize specific antigen followed by the immune response. In other words, all immunogens are antigens, but not all antigens are immunogens⁷². Therefore, vaccine candidate should be highly antigenic as well as immunogenic in nature. Consequently, the predicted vaccine construct was evaluated with VaxiJen 2.0 server⁴⁸ and IEDB Immunogenicity tool⁵² to ensure its antigenic and immunogenic potentials, respectively. On the contrary, allergenicity is the ability to create sensitization and allergic reactions associated with the IgE antibody response. So, vaccine candidate must be non-allergenic. We used AllerTOP 2.0 server to check the allergenicity of vaccine construct. Various physiochemical features of our vaccine protein *i.e.*, theoretical pI, instability index, in vitro and in vivo half-life, aliphatic index, molecular weight and grand average of hydropathicity (GRAVY) were assessed by using ProtParam server⁷³. Furthermore, the solubility of the vaccine protein upon overexpression in *E. coli* was predicted by SOLpro tool which has over 74% accuracy with 10-fold cross-validation⁴¹.

Prediction of secondary and tertiary structures. The primary sequence of vaccine protein was submitted to PSIPRED 4.0 server to predict the secondary structure. PSIPRED server uses two feed-forward neural networks which process the PSI-BLAST (Position-Specific Iterated-BLAST) generated a position-specific scoring matrix to predict a precise secondary structure from the input sequence⁷⁴. The normal Q₃ score of an earlier version (version 3.2) of PSIPRED server was 81.6%³⁴. The tertiary structure of a protein provides maximum stability in its lowest energy state through proper bending and twisting. Therefore, we used the RaptorX server to predict the 3D structure of our final vaccine construct⁷⁵. RaptorX server predicts 3D protein model based on multiple-template threading (MTT) and provides some confidence scores to indicate the quality of a predicted 3D model. For example, P-value for relative global quality, GDT, and uGDT for absolute global quality, and modelling error at each residue⁷⁵. The obtained 3D structure was visualized with BIOVIA Discovery Studio 2017.

Refinement and validation of 3D vaccine model. The crude model of our multiepitope vaccine protein was refined using GalaxyRefine web-server⁷⁶ which is based on the CASP10 tested refinement method⁷⁷. GalaxyRefine performs rehashed structure perturbation followed by overall structural relaxation by performing molecular dynamics simulation³⁵. Structure validation is an important step in homology modelling which works based on experimentally validated 3D protein structure. Hence, the refined vaccine model was submitted to ProSA-web for structural validation⁷⁸. ProSA calculates an overall quality score for a specific input structure. The quality score outside the characteristic range of native proteins indicates possible errors in the predicted protein structure. ERRAT server⁷⁹ was used to analyse the statistics of non-bonded interactions. Ramachandran plot was created using the RAMPAGE server⁸⁰. This server validates protein structure in terms of energetically allowed and disallowed dihedral angles psi (ψ) and phi (ϕ) of amino acid on the basis of PROCHECK principle⁸¹.

Prediction of B-cell epitopes in the vaccine protein. B lymphocyte provides humoral immunity by producing antibodies. Besides, they present antigen and secrete cytokines⁸². Therefore, the vaccine construct should have B-cell epitopes within their protein domains. As we mentioned earlier, B-cell epitopes can be divided into two categories: linear (continuous) and conformational (discontinuous). The presence of both linear and conformational B-cell epitopes was determined by using the ElliPro tool of IEDB server⁶³.

Disulphide engineering of the final vaccine. Disulphide engineering is a novel approach for creating disulphide bonds into the target protein structure. Disulphide bonds are covalent interactions that help in increasing the protein stability along with the examination of protein interactions and dynamics^{34,35}. Therefore, the refined model of the final vaccine construct was subjected to the Disulphide by Design 2.12³⁷ web-platform to perform disulphide engineering. Initially, the refined protein model was uploaded and run for the residue pair search that can be used for the disulphide engineering purpose. Potential residue pairs were selected for mutation with cysteine residue using create mutate function of the Disulfide by Design 2.12 server.

In silico codon adaptation and cloning. Codon adaptation is a way to increase the translational efficiency of foreign genes in the host when codon usage in both organisms differs from each other. In other words, unadapted codon may lead to the minor expression rate in the host. Therefore, the vaccine protein sequence was submitted to the Java Codon Adaptation Tool (JCAT)⁸³ to adapt their codon usage to the most sequenced prokaryotic organism, *E. coli* K12³⁴. Three additional options were selected to avoid the rho-independent transcription termination, prokaryote ribosome binding site, and restriction enzymes cleavage sites. CAI value⁸⁴ and GC content of the adapted sequence were calculated. Furthermore, XhoI and NdeI restriction sites were introduced to the N and C terminal of the sequence, respectively. Finally, the adapted nucleotide sequence corresponding to the designed vaccine construct was cloned into the *E. coli* pET28a(+) vector by using SnapGene 4.2 tool (<https://snapgene.com/>) to ensure the vaccine expression.

Molecular docking between the vaccine and TLR4 receptor. Molecular docking is a computational method which involves the interaction of a ligand molecule to the receptor molecules to give the stable adduct⁸⁵. It can also be used to predict the binding affinity between these two molecules in terms of scoring function. Toll-like receptor 4 (TLR4) is well-known for recognizing ligands from many Gram-negative bacteria. Besides, TLR4 activation may contribute to upregulation of TSPO/MBR related proteins through overproduction of cytokines⁶⁹. Therefore, TLR4 was selected as the receptor and obtained from the RCSB PDB database (PDB ID: 4G8A)⁸⁶ while the vaccine model was used as a ligand. Finally, the binding affinity between the multiepitope vaccine and TLR4 receptor was calculated through molecular docking using the ClusPro 2.0 server⁸⁷. This server completed the task in three successive steps such as rigid body docking, clustering of lowest energy structure, and structural refinement by energy minimization⁸⁷. The best docked complex was selected based on the lowest energy scoring and docking efficiency.

Dynamic simulations for vaccine stability. Molecular dynamic simulation (MDS) is essential to determine the stability of receptor-ligand complex³⁵. The MDS between TLR4 (as receptor) and vaccine (as ligand) was performed using AMBER14 force field⁸⁸ in YASARA Dynamics⁸⁹. Briefly, the complex was cleaned and optimized for hydrogen bond networks before the simulation followed by the creation of cubic simulation cell solvated with TIP3P water (density: 0.997 g/L⁻¹). A cut-off radius of 8.0 Å was used for short-range van der Waals and Coulomb interactions at physiological conditions (298 K, pH 7.4, 0.9% NaCl). Timestep of 1.25 fs was used and simulation snapshots were captured at every 100 ps. The MDS was executed for the period of 5 ns until the TLR4-Vaccine complex reaches a stable state. Finally, the trajectories generated from the simulation were analysed for the stability of the complex in terms of the root mean square deviation (RMSD) and root mean square fluctuation (RMSF) for backbone and side chain, respectively. Besides, comparison between initial and final protein backbones for hydrogen bonds was evaluated with BIOVIA Discovery Studio 2017.

Immune simulation for vaccine efficacy. *In silico* immune simulations were carried out using the C-ImmSim server (<http://150.146.2.1/C-IMMSIM/index.php>) to assess the immunogenic profile of multiepitope vaccine in real-life⁴⁴. C-ImmSim is an agent-based dynamic simulator for immune responses which uses position-specific scoring matrix (PSSM) and machine learning techniques for the prediction of epitopes and immune interactions, respectively⁴⁴. For most vaccines currently in use, the minimum recommended interval between dose 1 and dose 2 is 4 weeks⁹⁰. Therefore, three injections, containing 1000 vaccine proteins each, were administered (the parameters were set in C-ImmSim immune simulator) four weeks apart at 1, 84 and 168 time-steps (each time-step is equivalent to 8 hours in real-life and time-step 1 is injection at time = 0) with total 1050 simulation steps. All other simulation parameters were kept defaults. Furthermore, 12 injections of the designed peptide were given four weeks apart to simulate repeated exposure to the antigen seen in a typical endemic area so as to probe the clonal selection. The Simpson index, D (a measure of diversity) was interpreted from the plot.

Data availability. The data with particular importance are embedded within the submitted manuscript and the rests are provided as supplementary files. The sequence of proteins used in this study can be retrieved from IMG/M, RCSB PDB and NCBI database using their corresponding accession numbers.

Author Contributions

Z.N and U.K.A conceived and designed the study; Z.N and F.A carried out the major immunoinformatic analysis; Z.N and U.K.A wrote the manuscript; Z.N prepared the graphs and illustrations; Z.N, S.M.R.R, S.A.K, S.B.S, M.M.S, and Z.H performed the antigenicity and epitope predictions; M.M.K and M.M.R contributed to the critical revision of the manuscript; U.K.A supervised the whole work; and all authors approved the final manuscript.

Competing Interests

The author(s) declare no competing interests.

References

1. Breurec, S. *et al.* Genomic epidemiology and global diversity of the emerging bacterial pathogen *Elizabethkingia anophelis*. *Sci. Rep.* **6**, 30379 (2016). doi: 10.1038/srep30379.
2. Bulagonda, E. P. *et al.* Comparative genomic analysis of a naturally competent *Elizabethkingia anophelis* isolated from an eye infection. *Sci. Rep.* **8**, 8447 (2018). doi: 10.1038/s41598-018-26874-8.
3. Hu, S. *et al.* *Elizabethkingia anophelis* isolated from patients with multiple organ dysfunction syndrome and lower respiratory tract infection: report of two cases and literature review. *Front. Microbiol.* **8**, 382 (2017). doi: 10.3389/fmicb.2017.00382.
4. Coyle, A. L. *Elizabethkingia anophelis*: Exploring the outbreak of disease in the Midwest. *Nursing (Lond)*. **47**, 61–63 (2017). doi: 10.1097/01.nurse.0000512887.67622.84.
5. Li, Y. *et al.* Complete genome sequence and transcriptomic analysis of the novel pathogen *Elizabethkingia anophelis* in response to oxidative stress. *Genome Biol. Evol.* **7**, 1676–1685 (2015). doi: 10.1093/gbe/evv101.
6. Yung, C. F. *et al.* *Elizabethkingia anophelis* and association with tap water and handwashing, Singapore. *Emerg. Infect. Dis.* **24**, 1730–1733 (2018). doi: 10.3201/eid2409.171843.
7. Lin, J. N., Lai, C. H., Yang, C. H., Huang, Y. H. & Lin, H. H. Genomic features, phylogenetic relationships, and comparative genomics of *Elizabethkingia anophelis* strain EM361-97

- isolated in Taiwan. *Sci. Rep.* **7**, 14317 (2017). doi: 10.1038/s41598-017-14841-8.
8. Kampfer, P. *et al.* *Elizabethkingia anophelis* sp. nov., isolated from the midgut of the mosquito *Anopheles gambiae*. *Int. J. Syst. Evol. Microbiol.* **61**, 2670–2675 (2011). doi: 10.1099/ijs.0.026393-0.
9. Frank, T. *et al.* First case of *Elizabethkingia anophelis* meningitis in the Central African Republic. *Lancet* **381**, 1876 (2013). doi: 10.1016/S0140-6736(13)60318-9.
10. Teo, J. *et al.* First case of *E. anophelis* outbreak in an intensive-care unit. *Lancet* **382**, 855–856 (2013).
11. Lau, S. K. P. *et al.* *Elizabethkingia anophelis* bacteremia is associated with clinically significant infections and high mortality. *Sci. Rep.* **6**, 26045 (2016). doi: 10.1038/srep26045.
12. Castro, C. E. F. *et al.* *Elizabethkingia anophelis*: Clinical experience of an academic health system in Southeastern Wisconsin. *Open Forum Infect. Dis.* **4**, ofx251 (2017). doi: 10.1093/ofid/ofx251.
13. Nielsen, H. L. *et al.* Rare *Elizabethkingia anophelis* meningitis case in a Danish male. *JMM case reports* **5**, e005163 (2018). doi: 10.1099/jmmcr.0.005163.
14. Perrin, A. *et al.* Evolutionary dynamics and genomic features of the *Elizabethkingia anophelis* 2015 to 2016 Wisconsin outbreak strain. *Nat. Commun.* **8**, 15483 (2017). doi: 10.1038/ncomms15483.
15. Lau, S. K. P. *et al.* Evidence for *Elizabethkingia anophelis* transmission from mother to infant, Hong Kong. *Emerg. Infect. Dis.* **21**, 232–241 (2015). doi: 10.3201/eid2102.140623.
16. Li, W., Joshi, M. D., Singhanian, S., Ramsey, K. H. & Murthy, A. K. Peptide vaccine: progress and challenges. *Vaccines* **2**, 515–536 (2014). doi: 10.3390/vaccines2030515.
17. Jacob, C. O., Leitner, M., Zamir, A., Salomon, D. & Arnon, R. Priming immunization against cholera toxin and *E. coli* heat-labile toxin by a cholera toxin short peptide-beta-galactosidase hybrid synthesized in *E. coli*. *EMBO J.* **4**, 3339–3343 (1985).
18. Nandy, A., Basak, S., Nandy, A. & Basak, S. C. A brief review of computer-assisted approaches to rational design of peptide vaccines. *Int. J. Mol. Sci.* **17**, 666 (2016). doi: 10.3390/ijms17050666.
19. Kukutla, P. *et al.* Draft genome sequences of *Elizabethkingia anophelis* Strains R26T and Ag1 from the midgut of the malaria mosquito *Anopheles gambiae*. *Genome Announc.* **1**, (2013). doi: 10.1128/genomeA.01030-13.
20. Ali, M. *et al.* Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection. *Sci. Rep.* **7**, 9232 (2017). doi: 10.1038/s41598-017-09199-w
21. Ali, A. *et al.* Immunoinformatic and systems biology approaches to predict and validate peptide vaccines against Epstein–Barr virus (EBV). *Sci. Rep.* **9**, 720 (2019). doi: 10.1038/s41598-018-37070-z.
22. Mirza, M. U. *et al.* Towards peptide vaccines against Zika virus: Immunoinformatics combined with molecular dynamics simulations to predict antigenic epitopes of Zika viral proteins. *Sci. Rep.* **6**, 37313 (2016). doi: 10.1038/srep37313.
23. Ikram, A. *et al.* Exploring NS3/4A, NS5A and NS5B proteins to design conserved subunit multi-epitope vaccine against HCV utilizing immunoinformatics approaches. *Sci. Rep.* **8**, 16107 (2018). doi: 10.1038/s41598-018-34254-5.
24. Yasmin, T. & Nabi, A. H. B and T cell epitope-based peptides predicted from evolutionarily conserved and whole protein sequences of Ebola Virus as vaccine targets. *Scand. J. Immunol.* **83**, 321–337 (2016). doi: 10.1111/sji.12425.
25. Adhikari, U. K., Tayebi, M. & Rahman, M. M. Immunoinformatics approach for epitope-based peptide vaccine design and active site prediction against polyprotein of emerging

- Oropouche virus. *J. Immunol. Res.* **2018**, 1–22 (2018). doi: 10.1155/2018/6718083.
26. Adhikari, U. K. & Rahman, M. M. Overlapping CD8 + and CD4 + T-cell epitopes identification for the progression of epitope-based peptide vaccine from nucleocapsid and glycoprotein of emerging Rift Valley fever virus using immunoinformatics approach. *Infect. Genet. Evol.* **56**, 75–91 (2017). doi: 10.1016/j.meegid.2017.10.022.
27. Nezafat, N. *et al.* Designing an efficient multi-epitope peptide vaccine against *Vibrio cholerae* via combined immunoinformatics and protein interaction based approaches. *Comput. Biol. Chem.* **62**, 82–95 (2016). doi: 10.1016/j.compbiolchem.2016.04.006.
28. Droppa-Almeida, D., Franceschi, E. & Padilha, F. F. Immune-informatic analysis and design of peptide vaccine from multi-epitopes against *Corynebacterium pseudotuberculosis*. *Bioinform. Biol. Insights* **12**, 1177932218755337 (2018). doi: 10.1177/1177932218755337.
29. Lata, K. S. *et al.* Exploring Leptospiral proteomes to identify potential candidates for vaccine design against Leptospirosis using an immunoinformatics approach. *Sci. Rep.* **8**, 6935 (2018). doi: 10.1038/s41598-018-25281-3.
30. Saadi, M., Karkhah, A. & Nouri, H. R. Development of a multi-epitope peptide vaccine inducing robust T cell responses against brucellosis using immunoinformatics based approaches. *Infect. Genet. Evol.* **51**, 227–234 (2017). doi: 10.1016/j.meegid.2017.04.009.
31. Samykannu, G. *et al.* In silico characterization of B cell and T cell epitopes for subunit vaccine design of *Salmonella typhi* PgtE: a molecular dynamics simulation approach. *J. Comput. Biol.* **26**, 105–116 (2019). doi: 10.1089/cmb.2018.0010.
32. Khan, F., Srivastava, V. & Kumar, A. Computational identification and characterization of potential T-cell epitope for the utility of vaccine design against enterotoxigenic *Escherichia coli*. *Int. J. Pept. Res. Ther.* **25**, 289–302 (2019). doi: 10.1007/s10989-018-9671-3.
33. Hajighahramani, N. *et al.* Immunoinformatics analysis and in silico designing of a novel multi-epitope peptide vaccine against *Staphylococcus aureus*. *Infect. Genet. Evol.* **48**, 83–94 (2017). doi: 10.1016/j.meegid.2016.12.010.
34. Khatoon, N., Pandey, R. K. & Prajapati, V. K. Exploring Leishmania secretory proteins to design B and T cell multi-epitope subunit vaccine using immunoinformatics approach. *Sci. Rep.* **7**, 8285 (2017). doi: 10.1038/s41598-017-08842-w.
35. Pandey, R. K., Bhatt, T. K. & Prajapati, V. K. Novel Immunoinformatics approaches to design multi-epitope subunit vaccine for malaria by investigating Anopheles salivary protein. *Sci. Rep.* **8**, 1125 (2018). doi: 10.1038/s41598-018-19456-1.
36. Shey, R. A. *et al.* In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. *Sci. Rep.* **9**, 4409 (2019). doi: 10.1038/s41598-019-40833-x.
37. Craig, D. B. & Dombkowski, A. A. Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins. *BMC Bioinformatics* **14**, 346 (2013). doi: 10.1186/1471-2105-14-346.
38. Plotkin, S., Robinson, J. M., Cunningham, G., Iqbal, R. & Larsen, S. The complexity and cost of vaccine manufacturing - An overview. *Vaccine* **35**, 4064–4071 (2017). doi: 10.1016/j.vaccine.2017.06.003.
39. Busch, A. W. U. & Montgomery, B. L. The tryptophan-rich sensory protein (TSPO) is involved in stress-related and light-dependent processes in the Cyanobacterium *Fremyella diplosiphon*. *Front. Microbiol.* **6**, 1393 (2015). doi: 10.3389/fmicb.2015.01393.
40. Mathur, D. *et al.* PEPlife: a repository of the half-life of peptides. *Sci. Rep.* **6**, 36617 (2016). doi: 10.1038/srep36617.
41. Magnan, C. N., Randall, A. & Baldi, P. SOLpro: accurate sequence-based prediction of protein solubility. *Bioinformatics* **25**, 2200–2207 (2009). doi: 10.1093/bioinformatics/btp386.
42. Wlodawer, A. Stereochemistry and validation of macromolecular structures. *Methods Mol. Biol.* **1607**, 595–610 (2017). doi: 10.1007/978-1-4939-7000-1_24.

43. Duthie, M. S., Windish, H. P., Fox, C. B. & Reed, S. G. Use of defined TLR ligands as adjuvants within human vaccines. *Immunol. Rev.* **239**, 178–196 (2011). doi: 10.1111/j.1600-065X.2010.00978.x.
44. 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* **5**, e9862 (2010). doi: 10.1371/journal.pone.0009862.
45. Chen, I.-M. A. *et al.* IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res.* **47**, D666–D677 (2019). doi: 10.1093/nar/gky901.
46. Chen, I.-M. A. *et al.* IMG/M: integrated genome and metagenome comparative data analysis system. *Nucleic Acids Res.* **45**, D507–D516 (2017). doi: 10.1093/nar/gkw929.
47. Markowitz, V. M. *et al.* IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res.* **40**, D115–D122 (2012). doi: 10.1093/nar/gkr1044.
48. Doytchinova, I. A. & Flower, D. R. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics* **8**, 4 (2007). doi: 10.1186/1471-2105-8-4
49. Chaudhri, G. *et al.* T cell receptor sharing by cytotoxic T lymphocytes facilitates efficient virus control. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 14984 (2009). doi: 10.1073/pnas.0906554106.
50. Larsen, M. V. *et al.* Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics* **8**, 424 (2007). doi: 10.1186/1471-2105-8-424
51. Moutaftsi, M. *et al.* A consensus epitope prediction approach identifies the breadth of murine TCD8⁺-cell responses to vaccinia virus. *Nat. Biotechnol.* **24**, 817–819 (2006). doi: 10.1038/nbt1215
52. Calis, J. J. A. *et al.* Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS Comput. Biol.* **9**, e1003266 (2013). doi: 10.1371/journal.pcbi.1003266.
53. Dimitrov, I., Bangov, I., Flower, D. R. & Doytchinova, I. AllerTOP v.2 - a server for in silico prediction of allergens. *J. Mol. Model.* **20**, 2278 (2014). doi: 10.1007/s00894-014-2278-5.
54. Gupta, S. *et al.* In silico approach for predicting toxicity of peptides and proteins. *PLoS One* **8**, e73957 (2013). doi: 10.1371/journal.pone.0073957
55. Nielsen, M., Lundegaard, C. & Lund, O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinformatics* **8**, 238 (2007). doi: 10.1186/1471-2105-8-238
56. Wang, P. *et al.* Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC Bioinformatics* **11**, 568 (2010). doi: 10.1186/1471-2105-11-568.
57. Paul, S., Sidney, J., Sette, A. & Peters, B. TepiTool: A pipeline for computational prediction of T cell epitope candidates. *Curr. Protoc. Immunol.* **114**, 18.19.1–18.19.24 (2016). doi: 10.1002/cpim.12.
58. Luckheeram, R. V., Zhou, R., Verma, A. D. & Xia, B. CD4⁺ T cells: differentiation and functions. *Clin. Dev. Immunol.* **2012**, 1–12 (2012). doi: 10.1155/2012/925135.
59. Dhanda, S. K., Vir, P. & Raghava, G. P. Designing of interferon-gamma inducing MHC class-II binders. *Biol. Direct* **8**, 30 (2013). doi: 10.1186/1745-6150-8-30.
60. Dhanda, S. K., Gupta, S., Vir, P. & Raghava, G. P. S. Prediction of IL4 inducing peptides. *Clin. Dev. Immunol.* **2013**, 1–9 (2013). doi: 10.1155/2013/263952.
61. Nagpal, G. *et al.* Computer-aided designing of immunosuppressive peptides based on IL-10 inducing potential. *Sci. Rep.* **7**, 42851 (2017). doi: 10.1038/srep42851.
62. Cooper, M. D. The early history of B cells. *Nat. Rev. Immunol.* **15**, 191–197 (2015). doi: 10.1038/nri3801

63. Ponomarenko, J. *et al.* ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics* **9**, 514 (2008). doi: 10.1186/1471-2105-9-514.
64. Bui, H. H. *et al.* Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics* **7**, 153 (2006). doi: 10.1186/1471-2105-7-153.
65. Morris, G. M. *et al.* AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–2791 (2009). doi: 10.1002/jcc.21256.
66. Trott, O. & Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2009). doi: 10.1002/jcc.21334.
67. Lamiable, A. *et al.* PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. *Nucleic Acids Res.* **44**, W449–W454 (2016). doi: 10.1093/nar/gkw329.
68. Coffman, R. L., Sher, A. & Seder, R. A. Vaccine adjuvants: putting innate immunity to work. *Immunity* **33**, 492–503 (2010). doi: 10.1016/j.immuni.2010.10.002.
69. Wei, X. H. *et al.* The upregulation of translocator protein (18 kDa) promotes recovery from neuropathic pain in Rats. (2013). doi:10.1523/jneurosci.0324-12.2013.
70. Arai, R., Ueda, H., Kitayama, A., Kamiya, N. & Nagamune, T. Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Eng.* **14**, 529–532 (2001). doi: 10.1093/protein/14.8.529.
71. Gu, Y. *et al.* Vaccination with a paramyosin-based multi-epitope vaccine elicits significant protective immunity against *Trichinella spiralis* infection in Mice. *Front. Microbiol.* **8**, 1475 (2017). doi: 10.3389/fmicb.2017.01475.
72. Ilinskaya, A. N. & Dobrovolskaia, M. A. Understanding the immunogenicity and antigenicity of nanomaterials: Past, present and future. *Toxicol. Appl. Pharmacol.* **299**, 70–77 (2016). doi: 10.1016/j.taap.2016.01.005.
73. Wilkins, M. R. *et al.* Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol.* **112**, 531–552 (1999). doi: 10.1385/1-59259-890-0:571.
74. Buchan, D. W. A., Minneci, F., Nugent, T. C. O., Bryson, K. & Jones, D. T. Scalable web services for the PSIPRED protein analysis workbench. *Nucleic Acids Res.* **41**, W349–W357 (2013). doi: 10.1093/nar/gkt381.
75. Källberg, M. *et al.* Template-based protein structure modeling using the RaptorX web server. *Nat. Protoc.* **7**, 1511–1522 (2012). doi: 10.1038/nprot.2012.085.
76. Ko, J., Park, H., Heo, L. & Seok, C. GalaxyWEB server for protein structure prediction and refinement. *Nucleic Acids Res.* **40**, W294–W297 (2012). doi: 10.1093/nar/gks493.
77. Nugent, T., Cozzetto, D. & Jones, D. T. Evaluation of predictions in the CASP10 model refinement category. *Proteins* **82 Suppl 2**, 98–111 (2014). doi: 10.1002/prot.24377.
78. Wiederstein, M. & Sippl, M. J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* **35**, W407–W410 (2007). doi: 10.1093/nar/gkm290.
79. Colovos, C. & Yeates, T. O. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.* **2**, 1511–1519 (1993). doi: 10.1002/pro.5560020916.
80. Lovell, S. C. *et al.* Structure validation by Ca geometry: ϕ , ψ and C β deviation. *Proteins Struct. Funct. Bioinforma.* **50**, 437–450 (2003). doi: 10.1002/prot.10286.
81. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993). doi: 10.1107/S0021889892009944
82. Lund, F. E. Cytokine-producing B lymphocytes-key regulators of immunity. *Curr. Opin. Immunol.* **20**, 332–338 (2008). doi: 10.1016/j.coi.2008.03.003.

83. Grote, A. *et al.* JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res.* **33**, W526-W531 (2005). doi: 10.1093/nar/gki376.
84. Sharp, P. M. & Li, W. H. The codon adaptation index - a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* **15**, 1281–1295 (1987). doi: 10.1093/nar/15.3.1281.
85. Lengauer, T. & Rarey, M. Computational methods for biomolecular docking. *Curr. Opin. Struct. Biol.* **6**, 402–406 (1996). doi: 10.1016/S0959-440X(96)80061-3.
86. Berman, H. M. The protein data bank. *Nucleic Acids Res.* **28**, 235–242 (2000). doi: 10.1093/nar/28.1.235.
87. Kozakov, D. *et al.* The ClusPro web server for protein – protein docking. *Nat. Protoc.* **12**, 255–278 (2017). doi: 10.1038/nprot.2016.169.
88. Dickson, C. J. *et al.* Lipid14: the amber lipid force field. *J. Chem. Theory Comput.* **10**, 865–879 (2014).
89. Krieger, E., Darden, T., Nabuurs, S. B., Finkelstein, A. & Vriend, G. Making optimal use of empirical energy functions: Force-field parameterization in crystal space. *Proteins Struct. Funct. Bioinforma.* **57**, 678–683 (2004). doi: 10.1002/prot.20251.
90. Castiglione, F., Mantile, F., De Berardinis, P. & Prisco, A. How the interval between prime and boost injection affects the immune response in a computational model of the immune system. *Comput. Math. Methods Med.* **2012**, 1–9 (2012). doi: 10.1155/2012/842329.