2020 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS 2020), 20 June 2020, Selangor, Malaysia

Computational Investigation on Protein Sequence of Non-O157 VTEC for Potentiality of Vaccine Production

Md Fazlul Karim Khan
Faculty of Industrial Sciences & Technology
Universiti Malaysia Pahang
26300 Gambang, Kuantan, Pahang, Malaysia
stagnant_obit@yahoo.com

Muhammad Nomani Kabir*
Faculty of Computing
Universiti Malaysia Pahang
26300 Gambang, Kuantan, Pahang, Malaysia
nomanikabir@ump.edu.my
(corresponding author)

Shah Samiur Rashid
Faculty of Industrial Sciences & Technology
Universiti Malaysia Pahang
26300 Gambang, Kuantan, Pahang, Malaysia
samiur@ump.edu.my

Omar Tayan
Dept. of Computer Engineering & NOOR Research Center
College of Computer Science and Engineering
Taibah University
Madinah, Saudi Arabia
otayan@taibahu.edu.sa

Mohammad Nazmul Hasan Maziz
Graduate School of Medicine
Perdana University
Serdang, 43400 Selangor, Malaysia
poorpiku@yahoo.com

***Abstract*—Computational approach can be used for investigation of the protein sequences for developing a vaccine against infections. In this present study, a protein derived from non-O157 Verotoxin-producing *E. coli* (VTEC) was identified as a potential vaccine candidate that can be used to evaluate their immunogenicity and protective capability against VTEC infections. Identification of potential B-cell epitopes for promising vaccine, was carried out by evaluating with the methods of beta turns, hydropathicity, surface accessibility and antigenicity. The methods were implemented in MATLAB. However, our test results demonstrated that the VTEC-derived protein has plausible characteristics which provide significant insights for further investigations and will assist in relating potential drug targets/vaccine candidates.**

Keywords— Modeling, Protein sequencing, Verotoxin-producing E. coli, B-cell epitopes.

I. INTRODUCTION

Prediction and development of novel vaccine candidates by computational simulation approaches for numerous diseases have attained a positive response and greatly aided in the molecular synthesis to produce safe and effective vaccines. Diarrhea, one of the diseases, caused by Verotoxin-producing *E. coli* (VTEC) troubles the communities across the world [1]. A virulent group of bacteria called non-O157 VTEC is responsible for mild diarrhoea to severe bloody diarrhoea among the children and infants [2]. Treatment of infections becomes more complicated, and at times, almost unmanageable [3, 4]. In some cases, VTEC infection may lead to kidney failure and seldom death [5]. Virulent factors of non-O157 VTEC are vigorous and change with period and environment [1].

Understanding the molecular epidemiology of VTEC virulent factors is complex due to diversity and dynamic characteristics. Identification of a potential antigenic source to which an antibody binds, could be a key step towards decreasing the prevalence of VTEC pathogen as well as reducing the risk, prevent, or treat infections [6]. In principle, an antigen e.g., verotoxin in the form of a pathogen is neutralized through binding by antibodies produced from white blood cell which is known as B cell.

Efficient therapeutic antibodies and target specific vaccine development require the prediction of potential B-cell epitope which is the part of an antigen, recognized by antibodies generated by B cells [7]. The foremost reason for epitope prediction is to define an antigen in the potential antibody prediction, serological diagnosis and immunization. The prediction of B-cell epitopes plays an essential rule in antibody therapeutics, vaccines (peptide-based) and immunological diagnostic tools. Structural and functional epitope mapping methods are time-consuming and expensive and often fail to detect epitopes that are potentials. B-cell epitope prediction tools such as ABCpred, BepiPred, BCPreds and AAP algorithms are frequently used for antigenic portions of protein detection. However, successful identification and classification of proteins serve the importance of selecting suitable B-cell epitopes. Antigenicity, flexibility, beta-turn, surface accessibility and hydrophobicity are the main parameters to express a discontinuous and conformational B-cell epitope.

At present, there is no effective treatment or prevention for Hemolytic uremic syndrome (HUS) caused by VTEC [8]. This scenario provoked us to determine and classify potential B-cell epitopes by different parameters and scales of a computational approach for the expansion of novel immunogens. This approach examined encoded sequences which provides significant insights to discover potential vaccine candidates. Besides, B-cell epitopes identification may advance the therapeutics antibody against microbial toxins. Moreover, verotoxin B-cell with a high affinity to the antibody may introduce a new vaccine candidate against VTEC infections. This study aims to elucidate the potentiality of combining structural sequences and computational approaches in immunoinformatic analysis, to predict the

This research was supported by Universiti Malaysia Pahang (UMP) through University Research Grant (RDU1901150).

novel immunogens suitable vaccines candidate for the global prevention of infectious disease.

II. METHODOLOGY

In this research, a protein derived from non-O157 Verotoxin-producing *E. coli* (VTEC) was identified as a potential vaccine candidate. Non-O157 VTEC is clinically significant and was identified using microbiological assays. A protein Vtx1 derived from VTEC was confirmed by Sanger sequencing at Apical Scientific Sdn Bhd, Malaysia, and submitted to NCBI (National Center for Biotechnology Information) under the accession number BankIt2282323 MN696158. Amino acid sequence of Vtx1 is listed as follows:

LSCGGEGILPTSRTAKTYVDSLNVIRSAIGTPLQTISSG
GTSLLMIDSGTGDNLFAVDVRGIDPEEGRFNNLRLIVE
RNNLYVTGFVNRTNNVFYRFADFSHVTFPSRTTAVTL
SG

I. ALPHABETS AND THEIR CORRESPONDING SCALES IN KOLASKAR AND TONGAONKAR ANTIGENICITY AND CHOU AND FASMAN BETA TURN

Alphabet	Scale		Alphabet	Scale	
	Kolaskar and Tongaonkar antigenicity	Chou and Fasman beta turn		Kolaskar and Tongaonkar antigenicity	Chou and Fasman beta turn
A	1.064	0.66	M	0.826	0.6
C	1.412	1.19	N	0.776	1.56
D	0.866	1.46	P	1.064	1.52
E	0.851	0.74	Q	1.015	0.98
F	1.091	0.6	R	0.873	0.95
G	0.874	1.56	S	1.012	1.43
H	1.105	0.95	T	0.909	0.96
I	1.152	0.47	V	1.383	0.5
K	0.93	1.01	W	0.893	0.96
L	1.25	0.59	Y	1.161	1.14

II. NOMENCLATURE AND THEIR CORRESPONDING SCALES IN HYDROPATHICITY

Nomenclature	Scale	Nomenclature	Scale	Nomenclature	Scale
Ala	2.1	Gly	5.7	Pro	2.1
Arg	4.2	His	2.1	Ser	6.5
Asn	7	Ile	-8	Thr	5.2
Asp	10	Leu	-9.2	Trp	-10
Cys	1.4	Lys	5.7	Tyr	-1.9
Gln	6	Met	-4.2	Val	-3.7
Glu	7.8	Phe	-9.2		

Computational based recognition of epitopes by B-cells synthesizing vaccines induce most specific and an effective immune response which leads molecular synthesis to a significant label. However, currently most of the available prediction algorithms try to differentiate between epitopic and nonepitopic antigen surface residues [9]. The essential properties of predicting B-cell epitope based vaccine are flexibility, antigenicity, surface accessibility, hydrophilicity and epitope predictions [10]. We have analysed the amino acid sequence to predict a potential B-cell epitope by evaluating with the methods of beta turns, hydropathicity, surface accessibility and antigenicity by incorporating into MATLAB. The outcome of the result was validated through epitope-based server.

Fig. 1. provides a flowchart of the method for computing scores for a protein sequence. In this method, the potentiality of B-cell epitopes is derived in terms of score V using the inputs of protein sequence with size n, protein weight w and window size d. Scores are deduced by finding the average weights of the continuous epitopes of window size d. This process is done using the iterations with an inner loop k which progresses over the window size d. Note that each weight is added with s in each iteration, and thus, after all the iterations, we obtain the cumulative sum of the weights which is then averaged over d, proving the score V_i, where i is the position of the corresponding epitope. Weights are taken as the propensity scales of 20 amino acids as shown in Tables I-II are assigned based on the recommendation of different researchers [11]. Using the outer loop i over the whole sequence n, we obtain the scores V of all the epitopes.

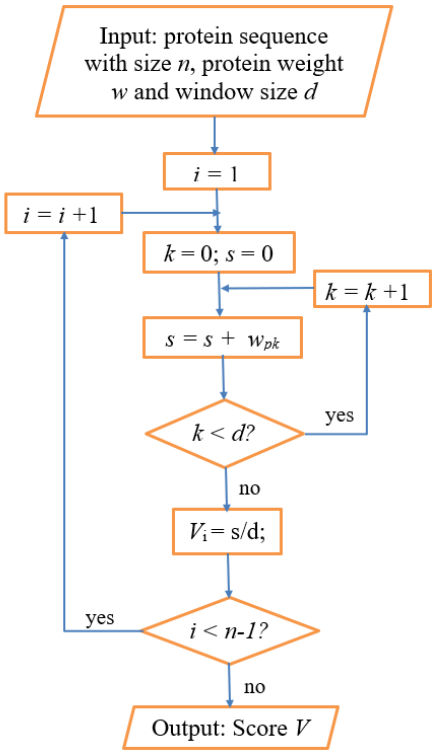

Fig. 1. Flowchart for computing scores for a protein sequence.

A. Chou and Fasman Beta-turn regions prediction

Beta turn region is used to identify epitope availability in a protein. A secondary structure element of protein consists of usually alpha-helix, beta-turn regions, and coil-coil regions. Chou and Fasman developed an algorithm of prediction scale for the secondary structure of proteins where beta-turn regions classify from alpha-helix and coil-coil regions. In this study, Chou and Fasman scales (Table 1) were used to compute the beta-turn regions [12] (Fig. 1).

The beta-turn value p(t) is determined as:

$$p(t) = f(j) \times f(j+1) \times f(j+2) \times f(j+3) \quad (1)$$

where *j* is the position of the amino acid in the four-residue window and *f*(*j*), *f*(*j*+1), *f*(*j*+2) and *f*(*j*+3) are bend frequencies in the four positions on the beta-turn.

B. *Hydropathicity*

Amino acid residues were defined using the hydropathicity scales proposed by Parker and colleagues [13]. These scales are commonly used for alpha-helices or membrane proteins prediction within the protein regions. It is an essential parameter to ensure the quality of drug design and delivery system. These hydropathicity scales (Table 2) were incorporated into the method in Fig. 2 to locate the surface residues by predicting the liner B-cell epitopes in the area of membrane proteins is in accordance with previous study[14].

C. *Surface accessibility*

Solvent accessibility of a protein implies the characteristics of protein structure and function. Surface probabilities determine amino acid to be hydrophilic on the surface of accessible protein [15]. The water-accessible surface (>20 A) is the most convenient surface area in protein. Surface accessibility is a parameter that influences antigenicity, while protein surfaces are equally accessible for antibody binding [16]. Surface accessibility of B cell epitopes is necessary, since hydrophilic regions are likely to initiate B cell immune response [10]. In this study, fractional surface probabilities of B-cell epitopes was determined according to Janin and colleagues algorithm [17]. Studies has reported that if surface area of a residue exceeds a threshold of 25%, is classified as exposed; however, there is no standard definition of the thresholds for solvent-accessible area [18].

D. *Antigenic prediction*

Antigenic determinants of proteins were identified by a semi-empirical method of Kolaskar and Tongaokar based on the physicochemical properties of amino acid residues occurrence frequency (Table I) in segmental epitopes and profusion in recognized B-cell epitopes [19]. This method calculates a large number of query proteins antigenicity with 75% accuracy while the development of computational methods provides the potential benefits for the reliable identification of B-cell epitopes [20].

E. *Probability of the epitopes based vaccines*

To validate the results, we assessed Vtx1 as a probable epitope based vaccine candidate through Vaxijen server (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html).

III. RESULTS AND DISCUSSION

Pathogenic isolates of non-O157 VTEC have a relatively high potential for developing resistance. Besides, potential genetic materials disseminate within the same species and as well as to bacteria in another genus or species [21]. Indeed, the flow of genetic materials between and within the genomes in bacteria is extensive and well established [22]. Computational approaches tools have a potentially powerful aid in vaccine development to induce effective immune responses. This approach is really useful, particularly for new or emerging pathogens when virulence factors/antigenic determinants knowledge is limited [20].

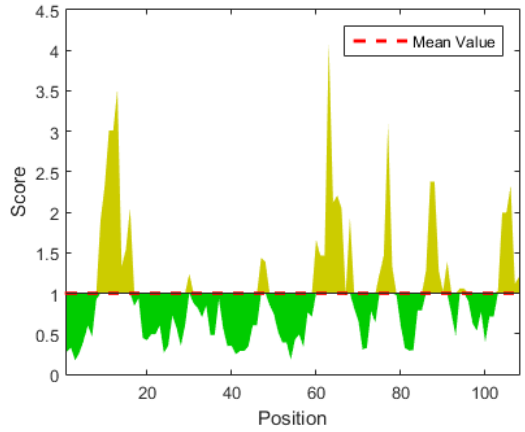

Fig. 2. Beta turn regions in a protein antigen for Vtx1.

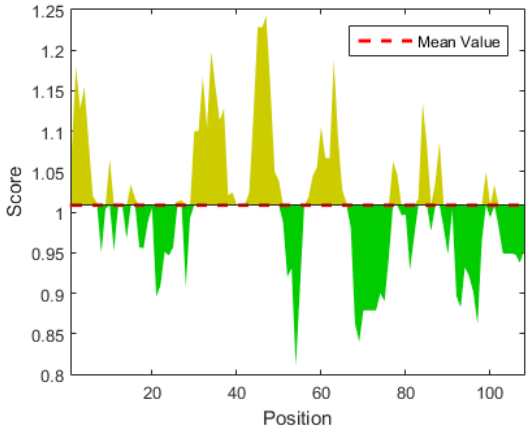

Fig. 3. Hydropathicity in antigen for Vtx1.

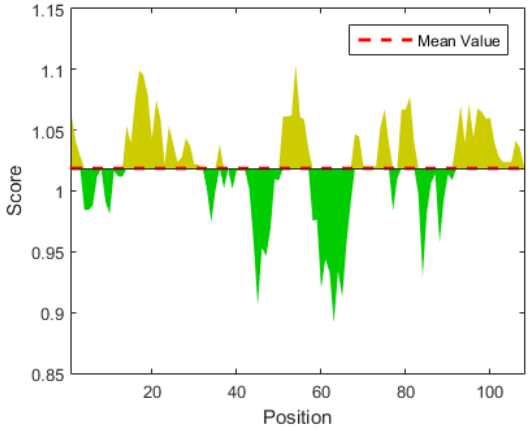

Fig. 4. Surface accessibility in antigen of Vtx1.

Epitope-based vaccines have notable advantage over the conventional ones[23] The most familiar computational methods of epitope prediction are sequence-based, and they specify linear or continuous B-cell epitopes [20]. Identifying a new B-cell epitope is one way of speed-up the expansion of novel vaccines. A noble vaccine candidate enhances the antigenic relationships among the hydrophilicity, surface accessibility, beta turns, exposed surface, the polarity of amino acids, which need to be recognized by immune systems

[24] and these features classify the potential B-cell epitope as a vaccine candidate against the non-O157 VTEC infections to avoid undesirable immune responses. Antigenicity and the ratio among B-epitopes numbers and the protein length need to be assessed to develop a vaccine [25]. However, a high antigenicity score and B-epitope density of proteins appear to be the best vaccine candidates [24].

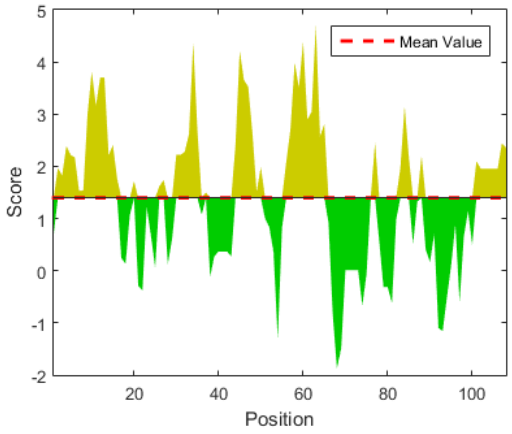

Fig. 5. Antigenicity in antigen of Vtx1.

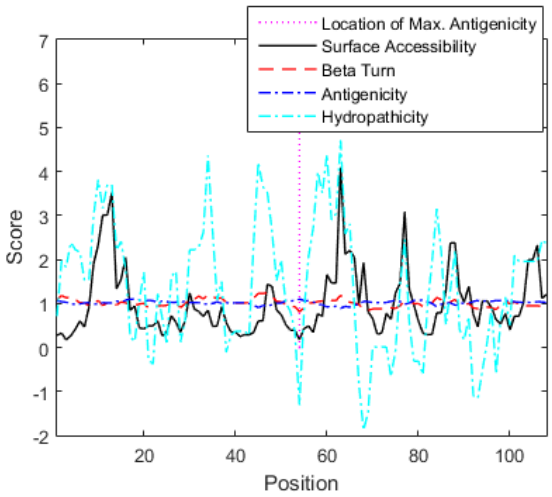

Fig. 6. Combined features for potential B-cell epitope of Vtx1.

The spikes above the threshold (residues are colored in yellow on Fig. 2) are probable beta-turn regions. The probable beta-turn region of protein Vtx1(48.64%) was classified from alpha-helix and coil-coil regions at the threshold setting of 1.0 is an agreement with Mtimet et al. [26]. Studies have revealed that the antigenic portion is located in beta-turn regions of a protein [10]. Besides, this study reveals that Vtx1 (45.94%) epitope was hydrophilic at a threshold setting of 1.01 (Fig. 3) is in accordance with Manee et al. [27]. The most accessible region of B-cell epitopes in Vtx1 (33.33%) was classified at a threshold setting of 1.02 (Fig. 4) is an agreement with Fanuel et al. [14]. Last, the antigenicity of Vtx1(51.35%) epitope was recorded at a threshold setting of 1.2, may serve as possible targets for the construction of epitope-based diagnostics and vaccines (Fig. 5) whereas a recent study calculated the antigenicity based on threshold score of 1.00 [14]. A recent study has documented that 0.4 thresholds reflect as a good antigenic epitope [28].

The graphical representations correspond to the positions of the residues in Vtx1 protein. The higher score (averaged to a specified window) for the residues interprets as a higher probability of an epitope.

Besides, the most antigenic region of Vtx1 was classified by surface accessibility, beta-turn, antigenicity and hydropathicity for a standard threshold to find the potential antigenic epitopes (Table III) which is in an agreement with previous study [29]. Vtx1 has revealed a score of Vtx1 0.4509 at cut-off 0.4, which considers as a probable antigen and its agrees with several recent studies [30, 31]

III. VTEC Protein Residue With A Higher Probability For Potential Epitope

Proteins	Estimation	Surface Accessibility	Beta Turn	Antigenicity	Hydropathicity
VTX1	Threshold value	1.02	1	1.2	1.01
	Probability (%)	33.33	48.64	51.35	45.94

A graphical representation of all the features (surface accessibility, beta-turn, antigenicity, and hydropathicity) is presented to specify a potential B-cell epitopes region in protein antigen Vtx1 (Fig. 6).

This study revealed that in Vtx1 protein, DSGTGDNLF (47-55) and MIDSGTGDN (45-53), two 9-mer peptides with a score of 1.242 and 1.229, respectively could be the most effective B-cell epitopes. This potential B-cell epitope regions which are highly antigenic, hydrophilic, surface accessible and situated on the beta-turn regions can be of a great interest in understanding disease etiology and assays for diagnosis and developing epitope-based vaccines.

IV. Conclusion

Existence, spread and prevalence of verotoxin producing non-O157 *E. coli* pose a risk to public health. A systematic screening policy should be implemented for conformational B-cell epitope prediction with an updated dataset in a simple, inexpensive and effective way. The current study serves the purpose of prediction for potential antigenic epitopes and suggest that VTEC-derived Vtx1 protein has a potential B-cell epitope for the formulation of both human and veterinary vaccines for adoptive immunotherapy against VTEC infections. Besides, prediction of B-cell epitopes using computational approaches will attribute to cost and time effective which enhance the accessibility for epitope prediction and implications on vaccine development. Further studies can be carried out by incorporating optimization algorithms [32, 33] for better prediction of B-cell epitopes.

Acknowledgment

We are grateful to Universiti Malaysia Pahang, Hospital Tengku Ampuan Afzan, Malaysia and Medical Research & Ethics Committee (MREC), Malaysia, for research facilities and ethical approve.

REFERENCES

[1] C. Conrad, K. Stanford, T. McAllister, J. Thomas, and T. Reuter, "Shiga toxin-producing *Escherichia coli* and current trends in diagnostics," (in English), *Animal Frontiers*, vol. 6, no. 2, pp. 37-43, Apr 2016.

[2] V. Peirano, M. N. Bianco, A. Navarro, F. Schelotto, and G. Varela, "Diarrheagenic *Escherichia coli* associated with acute gastroenteritis in children from Soriano, Uruguay," *Can J Infect Dis Med Microbiol*, vol. 2018, p. 8387218, 2018.

[3] M. Nazmul, M. Fazlul, and M. Rashid, "Plasmid profile analysis of non-O157 diarrheagenic *Escherichia coli* in Malaysia," *Indian Journal of Science*, vol. 1, no. 2, pp. 130-132, 2012.

[4] M. Fazlul, S. S. Rashid, M. Nazmul, I. Zaidul, R. Baharudin, and A. Nor, "A clinical update on antibiotic resistance gram-negative bacteria in Malaysia-A review," *Journal of International Pharmaceutical Research*, vol. 45, pp. 270-283, 2018.

[5] A. Y. Tamime, *Microbial toxins in dairy products*. John Wiley & Sons, 2017.

[6] S. F. Hussin, "The Detection Eschericha Coli Bacteria: A Review Of Image Processing Methods," *International Journal of Software Engineering and Computer Systems*, vol. 5, no. 2, pp. 26-36, 2019.

[7] N. Zobayer, A. A. Hossain, and M. A. Rahman, "A combined view of B-cell epitope features in antigens," *Bioinformation*, vol. 15, no. 7, pp. 530-534, 2019.

[8] M. Golshani, M. Oloomi, and S. Bouzari, "In silico analysis of Shiga toxins (Stxs) to identify new potential vaccine targets for Shiga toxin-producing *Escherichia coli*," (in eng), *In Silico Pharmacol*, vol. 5, no. 1, p. 2, Dec 2016.

[9] V. Demolombe, A. G. de Brevern, F. Molina, G. Lavigne, C. Granier, and V. Moreau, "Benchmarking the PEPOP methods for mimicking discontinuous epitopes," *BMC Bioinformatics*, vol. 20, no. 1, p. 738, 2019/12/30 2019.

[10] M. A. Khan, M. U. Hossain, S. M. Rakib-Uz-Zaman, and M. N. Morshed, "Epitope-based peptide vaccine design and target site depiction against Ebola viruses: an immunoinformatics study," *Scand J Immunol*, vol. 82, no. 1, pp. 25-34, Jul 2015.

[11] S. Moelbert, E. Emberly, and C. Tang, "Correlation between sequence hydrophobicity and surface-exposure pattern of database proteins," *Protein Science*, vol. 13, no. 3, pp. 752-762, 2004.

[12] P. Y. Chou and G. D. Fasman, "Prediction of the secondary structure of proteins from their amino acid sequence," (in eng), *Adv Enzymol Relat Areas Mol Biol*, vol. 47, pp. 45-148, 1978.

[13] J. M. Parker, D. Guo, and R. S. Hodges, "New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites," (in eng), *Biochemistry*, vol. 25, no. 19, pp. 5425-32, Sep 23 1986.

[14] S. Fanuel, S. Tabesh, E. Sadroddiny, and G. A. Kardar, "Analysis of predicted B and T-cell epitopes in Der p 23, allergen from dermatophagoides pteronyssinus," (in eng), *Bioinformation*, vol. 13, no. 9, pp. 307-312, 2017.

[15] J. Novotny and E. Haber, "Static accessibility model of protein antigenicity: the case of scorpion neurotoxin," *Biochemistry*, vol. 25, no. 22, pp. 6748-54, Nov 4 1986.

[16] J. Novotny *et al.*, "Antigenic determinants in proteins coincide with surface regions accessible to large probes (antibody domains)," (in eng), *Proc Natl Acad Sci U S A*, vol. 83, no. 2, pp. 226-30, Jan 1986.

[17] J. Janin, S. Miller, and C. Chothia, "Surface, subunit interfaces and interior of oligomeric proteins," (in eng), *J Mol Biol*, vol. 204, no. 1, pp. 155-64, Nov 5 1988.

[18] B. Zhang, L. Li, and Q. Lü, "Protein solvent-accessibility prediction by a stacked deep bidirectional recurrent neural network," (in eng), *Biomolecules*, vol. 8, no. 2, p. 33, 2018.

[19] A. S. Kolaskar and P. C. Tongaonkar, "A semi-empirical method for prediction of antigenic determinants on protein antigens," *FEBS Lett*, vol. 276, no. 1-2, pp. 172-4, Dec 10 1990.

[20] L. Liljeroos, E. Malito, I. Ferlenghi, and M. J. Bottomley, "Structural and computational biology in the design of immunogenic vaccine antigens," *Journal of Immunology Research*, vol. 2015, 2015.

[21] M. Gull and S. El-Baz, "Introductory Chapter: Preface to Plasmids," 2018.

[22] R. Pinilla-Redondo, V. Cyriaque, S. Jacquiod, S. J. Sorensen, and L. Riber, "Monitoring plasmid-mediated horizontal gene transfer in microbiomes: Recent advances and future perspectives," *Plasmid*, vol. 99, pp. 56-67, Sep 2018.

[23] T. A. Ahmad, A. E. Eweida, and S. A. Sheweita, "B-cell epitope mapping for the design of vaccines and effective diagnostics," *Trials in Vaccinology*, vol. 5, pp. 71-83, 2016/01/01/ 2016.

[24] M. Meunier, M. Guyard-Nicodeme, E. Hirchaud, A. Parra, M. Chemaly, and D. Dory, "Identification of Novel Vaccine Candidates against Campylobacter through Reverse Vaccinology," (in English), *Journal of Immunology Research*, vol. 2016, 2016.

[25] M. Oprea and F. Antohe, "Reverse-vaccinology strategy for designing T-cell epitope candidates for *Staphylococcus aureus* endocarditis vaccine," *Biologicals*, vol. 41, no. 3, pp. 148-53, May 2013.

[26] G. Mtimet, M. Stayoussef, and B. Yacoubi-Loueslati, "Prediction of the most probable B Cell epitopes from (DnaK) adhesin of mycobacterium tuberculosis Using Immunoinformatic tools," *International Journal of Peptide Research and Therapeutics*, vol. 26, no. 1, pp. 477-485, 2020/03/01 2020.

[27] M. M. Manee *et al.*, "Molecular cloning, bioinformatics analysis, and expression of small heat shock protein beta-1 from camelus dromedarius, Arabian camel," *PloS One*, vol. 12, no. 12, 2017.

[28] F. Bande, S. S. Arshad, M. Hair Bejo, S. Kadkhodaei, and A. R. Omar, "Prediction and in silico identification of novel b-cells and t-cells epitopes in the s1-spike glycoprotein of M41 and CR88 (793/B) infectious bronchitis virus serotypes for application in peptide vaccines," (in eng), *Advances in Bioinformatics*, vol. 2016, pp. 5484972-5484972, 2016.

[29] N. Palanisamy and J. Lennerstrand, "Computational prediction of Usutu virus E protein B cell and T cell epitopes for potential vaccine development," *Scandinavian Journal of Immunology*, vol. 85, no. 5, pp. 350-364, 2017.

[30] B. D. Sanasam and S. Kumar, "PRE-binding protein of plasmodium falciparum is a potential candidate for vaccine design and development: An in silico evaluation of the hypothesis," *Medical Hypotheses*, vol. 125, pp. 119-123, 2019.

[31] V. Subramaniyan, R. Venkatachalam, P. Srinivasan, and M. Palani, "In silico prediction of monovalent and chimeric tetravalent vaccines for prevention and treatment of dengue fever," *Journal of Biomedical Research*, vol. 32, no. 3, p. 222, 2018.

[32] J. B. Odili and M. N. M. Kahar, "African buffalo optimization," *International Journal of Software Engineering and Computer Systems*, vol. 2, no. 1, pp. 28-50, 2016.

[33] M. N. Kabir, J. Ali, A. A. Alsewari, and K. Z. Zamli, "An adaptive flower pollination algorithm for software test suite minimization," in *2017 3rd International Conference on Electrical Information and Communication Technology (EICT)*, 2017, pp. 1-5: IEEE.