

STRUCTURAL ANALYSIS OF PLASMID-MEDIATED VEROTOXIN GENE PRODUCING NON-O157 *ESCHERICHIA COLI* USING MOLECULAR DYNAMIC SIMULATIONS

Md Fazlul Karim Khan¹, Shah Samiur Rashid^{*1}, Aizi Nor Mazila Ramli¹, Muhammad Nomani Kabir², and Mohammad Nazmul Hasan Maziz³

Address(es):

¹Faculty of Industrial Sciences & Technology, Universiti Malaysia Pahang, Gambang, 26300 Pahang, Malaysia.

²Faculty of Computing, Universiti Malaysia Pahang, Gambang, 26300 Pahang, Malaysia.

³Graduate School of Medicine, Perdana University, Wisma Chase Perdana, Changkat Semantan, Damansara Heights, 50490 Kuala Lumpur, Malaysia.

*Corresponding author: samiur@ump.edu.my, stagnant_obit@yahoo.com

<https://doi.org/10.15414/jmbfs.4142>

ARTICLE INFO

Received 28. 12. 2020

Revised 25. 2. 2021

Accepted 3. 3. 2021

Published 1. 8. 2021

Regular article



ABSTRACT

Plasmid-mediated Non-O157 VTEC is a pathogenic *E. coli* serotype, are responsible for many life-threatening diseases such as diarrhoea. The thermostable crystal structure has been widely sought after for industry and therapeutic applications through the structural analysis. The National Center for Biotechnology Information (NCBI) Genbank Database has been sourced to obtain plasmid-mediated Verotoxin genes producing non-O157 MN 696158 (Vtx1-1) and MN688720 (Vtx2) sequences. The tertiary structure of MN696158 (Vtx1-1) and MN688720 (Vtx2) was generated by operating MODELLER. The result exposed multiple templates during modelling processes have improved the local stereochemical quality of the produced models. The structural analysis also disclosed the similarities and differences between the models (Vtx1-1 and Vtx2). Furthermore, the thermal stability profile of plasmid-mediated non-O157 VTEC were studied. The molecular dynamics simulations of plasmid-mediated non-O157 VTEC structures (Vtx1-1 and Vtx2) illustrated the interactions between amino acids. On the other hand, Vtx1-1 and Vtx2 showed noticeable differences in their relative conformational flexibility and stability at elevated temperature. However, it is expected that the information of the thermal stable of plasmid-mediated non-O157 VTEC models can be used for potential vaccine candidate through protein engineering in future.

Keywords: Verotoxin gene, GROMACS, Molecular dynamic simulations, Non-O157 *E. coli*

INTRODUCTION

Non-O157 Verotoxin producing *E. coli* (VTEC) is gradually recognised as a critical enteric foodborne pathogen which associates with life threatening complications (Colello *et al.*, 2016; Khan *et al.*, 2020). The production of verotoxin related virulence properties causes mild diarrhoea, hemorrhagic colitis (HC), and fatal hemolytic uremic syndrome (HUS) across the worldwide. Generally, *E. coli* O157: H7 serotypes associated with foodborne illness and non-O157 serotypes have been accused of gastroenteritis and HUS outbreaks (Public Health England, 2018). Furthermore, VTEC has become a major public health issue in combination with serious food-borne disease and diseases (Mei *et al.*, 2015; Khan *et al.*, 2020).

In eukaryotic cells, Vtx is one of the virulent factors which produces one or more toxin which. inhibits protein synthesis. (Cheung & Trachtman, 2014). These are known as Shiga toxin-producing *E. coli* (STEC) or Verotoxin producing *E. coli* (VTEC) (Nazmul *et al.*, 2012; Kaper & O'Brien, 2014). The Vtx genes (Vtx1 and Vtx2) have the same structure and variation of the Vtx 1 and Vtx 2 sequences and variants has been defined (Hazards, 2013; Khan *et al.*, 2020). Moreover, Vtx is the usual member of the bacteria of type 2 ribosome-inactivating proteins (RIPs) (Zhu *et al.*, 2018). In laboratory conditions, the determination of the protein structure requires time and energy and is not an economic process. (Hai-You *et al.*, 2016). In knowledge-based material, structural analysis of proteins is crucial while 3D (3-D) arrangement of amino acid atoms could provide the best solution for homology or comparative modelling. The simulation of molecular dynamics can also improve structure prediction with the lowest error and loss of data (Geng *et al.*, 2019). X-ray crystallography and NMR is typically the most effective approach to solve protein structures across various steps. (Su *et al.*, 2015). Protein structures of the same molecule have four distinct stages, referring to the prediction of secondary protein structure, modelling homology, molecular docking and molecular dynamic simulations. (Muhammed & Aki-Yalcin, 2019).

Pathogenic associations with host cells also affect infectious diseases in practise (Fazlul *et al.*, 2011; Fazlul *et al.*, 2018). Thus, we attempt to generate multi-

templates of plasmid-mediated verotoxin models. The truth is that a single template cannot cover the entire verotoxin sequence and thus multiple templates should be implemented to cover the entire sequence of the modelled verotoxin. In addition, a multiple templates combination may also cover one template's weakness by another template (Chakravarty *et al.*, 2008). The models created allows the analysis of interactions in plasmid-mediated non-O157 *E. coli* verotoxin gene sequences. The predicted verotoxin structures were also simulated to show conformation changes in the height temperature. This provides explicit structural information to the verotoxin gene that is essential for better enzyme control.

METHODS AND MATERIALS

Model development

The plasmid mediated verotoxin producing amino acid sequences of MN696158 (Vtx1-1) and MN688720 (Vtx2) were obtained from NCBI Genbank and BLAST for finding acceptable models against the PDB database. Single and multiple models that have been chosen for Modelling was based on the coverage length, identity of the sequence and differences between the target and the template. Plasmid mediated verotoxin gene via MODELLER version 9.20 (Heo & Feig, 2018), models were produced. A hundred suitable models based on the lowest discrete optimised protein energy (DOPE) scores were chosen for the best one. The appropriate model was then subjected to a molecular dynamics simulation model refinement process.

Refinement process

The refinement approach was implemented with some alterations (Heo & Feig, 2018). First of all, the local stereochemistry of the chosen component models of verotoxin were refined using locPREFMD server before simulation of molecular dynamics using 5.1.4 GROMACS (Heo & Feig, 2018). The simulation was done by in a simple cubic box with a minimum of 10 Å from the edge of the box with

a force field of CHARMM36m (Huang *et al.*, 2017). With TIP3P water molecules, the mechanism was solved, and a sufficient number of sodium ions was neutralized. The ensemble was after that, energy reduced by 5000 steps of the steepest descent algorithm and balanced at NVT phase to 298 K (constant number of particles, volume and temperature) and 1 bar pressure at 1 bar pressure NPT process for 100 ps, respectively (constant number of particles, pressure and temperature). Ultimately, the simulation was performed for 50 ns at 298 K and 1 bar pressure. To constrain the bond, LINear Constraint Solver (LINCS) was used, whereas the electrostatic interactions were assessed by both Coulomb and van der Waals interactions, the particle mesh Ewald process (PME) with a 12 Å cut-off. The integration time stage was 2 fs, and every 1 ps during the simulation, a snapshot of the structures was taken. The stability of the trajectory was assessed using the GROMACS functionality by root-mean-square deviation (RMSD) and radius of gyration (Rg). For the final round of local stereochemistry refinement, the stabilized structure was extracted and introduced to locPREFMd.

Model evaluations

The stereochemical content of the VTX models was determined and evaluated by Verify 3D (Chakkyarath & Natarajan, 2019), ERRAT (Soumiya *et al.*, 2019) and PROCHECK (Chakkyarath & Natarajan, 2019) via structure analysis and verification server version 5.0 (SAVES 5.0). The energy level of the VTX models was calculated using Swiss-PdbViewer (Amir *et al.*, 2019). The molecular visualization of the protein-protein interactions was demonstrated in the PyMOL 2.2.2 edition (Faure *et al.*, 2019).

Thermal study for protein deterioration

Protein deterioration analysis plays an essential role in the process of thermal stability determination with exponential growth or decay with time. The Vtx sample was prepared using proteolytic activity assays as described by Duanisassaf (Duanis-Assaf *et al.*, 2020) to determine the optimum temperature and allowed to hydrolyse casein for 10 minutes at 20 °C, 30 °C, 40 °C, 50 °C and 60 °C. The reaction was terminated by trichloroacetic acid (TCA) (concentration of 110 mM), and the reaction mixture was centrifuged, and the supernatant was mixed with Na₂CO₃ (concentration of 500 mM) while F-C phenol reagent (concentration of 0.5 M) for colour development. In the preparation of (TCA) and

Folin and Ciocalteu’s (F-C) phenol, both the reagents were diluted to a final concentration of 110 mM and 0.5 M, respectively from their stock solution. Sodium carbonate (Na₂CO₃) buffer was prepared by adding 53 g of anhydrous Na₂CO₃ in 1 L of distilled water to a final concentration of 500 mM. After that, the crude samples were incubated at five elevated temperatures of 20 °C, 30 °C, 40 °C, 50 °C, and 60 °C without substrate for 1 hour. Similarly, the hydrolysis was conducted at 37 °C, as described above. A plot of Vtx activity against elevated temperatures was generated.

Thermal dynamic behaviours

The refined verotoxin models were set to the molecular dynamic simulation to examine dynamics and conformation variations at the specified temperature. Three separate systems have been developed for each verotoxin model, dissolved with TIP3P water molecules and neutralized by a sufficient number of sodium ions. After that, energy was minimized by 5000 steps of the steepest descent algorithm and balanced at 300 K, 313 K and 323 K, respectively, through NVT and NPT. Consequently, the simulation was done for 100 ns. LINCS was used to constrain bond length, while electrostatic interactions were assessed by SMEs with a 12 Å cut-off for both Coulomb and van der Waals correlations. The integration time stage was 2 fs and the structure snapshot was reported every 1 ps during the simulation. Besides, dynamic behavior and conformational changes of VTX models were investigated by RMSD analysis, root-mean-square fluctuation (RMSF), Rg, solvent-accessible surface area (SASA) and hydrogen bond number.

RESULTS AND DISCUSSION

Comparative modelling and model refinement

The similarities and differences among the Vtx1-1 and Vtx2 were determined while comparison studies disclosed important structural information of Vtx protein. The isolated Vtx gene sequences were BLAST against the PDB database for an appropriate Vtx protein model. The protein BLAST result revealed that Vtx1-1 are vastly identical to 1R4Q_A with percent identity of 95.11 %, while 96.21 % identity to 4M1U_A was observed in Vtx2 (Table 1 and Table 2).

Table 1 Templates of protein revealed the maximum similarity with Vtx1-1 sequences based on various parameters

Template	PDB report	Query analysis (%)	E-value	Identity (%)	Gaps (%)	Score
4M1U_A	Chain A, Shiga toxin	83	2e-77	55.61	10	209
1R4P_A	Chain A, Shiga toxin	83	2e-67	57.61	10	209
1DMO_A	Chain A, Shiga toxin	88	5e-118	95.05	10	337
1R4Q_A	Chain A, Shiga Toxin	88	4e-118	95.11	10	337
4P2C_A	Chain A, Shiga toxin	83	3e-66	57.58	10	206

Table 2 Templates of protein revealed the maximum similarity with Vtx2 sequences based on various parameters

Template	PDB report	Query analysis (%)	E-value	Identity (%)	Gaps (%)	Score
4M1U_A	Chain A, Shiga toxin	100	7e-147	96.21	3	400
1R4P_A	Chain A, Shiga toxin	100	7e-147	96.21	3	400
4P2C_A	Chain A, Shiga Toxin	100	9e-135	88.63	3	369
1R4Q_A	Chain A, Shiga toxin	98	9e-90	61.06	4	255
1DMO_A	Chain A, Shiga toxin	11	1.6	26.09	0	16.9

However, the query coverage of Vtx1-1 by 1R4Q_A was only 88 %, although Vtx2 exhibited the most extended query coverage of 100 %. Furthermore, the least E-value and gap with 4M1U_A, 1R4Q_A, 4P2C_A, 1DMO_A, and 1R4P_A was considered to finalize based on the recommendation by several studies (Kerfeld & Scott, 2011; Frith, 2019) for an appropriate alignment amongst the sequences.

Primarily, 1R4Q_A has appeared as the best template due to the most extended coverage (88 %) and sequence identity of 95.11 % for Vtx1-1 protein model. Besides, 4M1U_A has the highest coverage (100 %) with the sequence identity of 96.21 % and was selected for Vtx2. According to a recent study, protein modelled with a template of 30-50 % identity will not exceed 4 Å RMSD from its native structure (Monzon *et al.*, 2017). Furthermore, the gap between gene sequences has to very low (1-2 %) to evade misalignment amongst the target and template sequence, which may mislead the structured protein (Dorn *et al.*, 2014). The protein sequences of Vtx1, Vtx1-1, and Vtx2 was aligned with the available protein data structure in NCBI databases associated with the PDB database. The most compatible protein data was selected based on the various parameters to build up the perfect models.

In this study, a set of 100 templates for each Vtx sequences was generated to provide a pool of perfect results (Sefidbakht *et al.*, 2017). The excellent model was distinguished from the bad models using the MODELLER in-built assessment method based on DOPE score. DOPE is a potential statistical assessment that corresponds to non-interacting atoms in a homogeneous sphere with the radius dependent on a native sample structure (Jing & Dong, 2017). A

lower DOPE score indicates the structure has a better packing of the atoms and is more accurate at its native conformation (Cloete *et al.*, 2018). These scoring methods are precise sufficiently to select the most authentic model among the generated models using the MODELLER (Webb & Sali, 2017).

The stereochemical consistency of the desired models must be carried out for further assessment. These selected models were investigated using the SAVES 5.0 server using various evaluation programs such as Verify 3D, which calculates the number of residues in an atomic model consistent with its amino acid sequence of 80 % (Tran *et al.*, 2015). Meanwhile, ERRAT analysed the statistics of non-bonded interactions amongst various atom types (Wei *et al.*, 2017); and PROCHECK assessed the protein stereochemical quality of a protein models in favoured, allowed and outlier regions (Elangoe *et al.*, 2014). These scoring methods are often satisfactorily adequate to select the most accurate generated models (Haddad *et al.*, 2020).

Among the accurate generated models, 4M1U_A for Vtx1-1 protein was selected based on the scoring methods shown in Table 3. However, the Vtx1-1 protein structure passed the Verify 3D test with a score of 93.82 %, ERRAT (97.76) and PROCHECK (93.4 %). Meanwhile, 4P2C_A was selected template for Vtx2 protein with a Verify 3D score of 94.95 %, ERRAT (87.28) and PROCHECK (87.6 %) have satisfactory phi and psi dihedral angle allocations of amino acid residues in the modelled structures. Both Vtx1-1 (4M1U_A) and Vtx2 (4P2C_A) revealed an excellent quality of predicted protein model based on the evaluation tool ProQ server (Sajib *et al.*, 2020).

Table 3 Models assessment using Verify 3D, ERRAT and PROCHECK

Model	Template	Verify 3D (%)	ERRAT (%)	PROCEHCK (%)
Vtx1-1	4M1U_A	93.82	97.76	Favoured region:93.4 Allowed region:6.6 Outlier region:0.0
	1R4P_A	93.98	95.29	Favoured region:92.6 Allowed region:7.4 Outlier region:0.0
	1DM0_A	75.49	100	Favoured region:85.3 Allowed region:14.9 Outlier region:0.0
	1R4Q_A	87.29	91.92	Favoured region:85.9 Allowed region:14.0 Outlier region:0.1
	4P2C_A	94.95	87.28	Favoured region:87.6 Allowed region:11.7 Outlier region:0.8
Vtx2	4M1U_A	87.29	91.92	Favoured region:85.9 Allowed region:14.0 Outlier region:0.1
	1R4P_A	92.15	83.18	Favoured region:92.6 Allowed region:7.4 Outlier region:0.0
	4P2C_A	94.95	87.28	Favoured region:87.6 Allowed region:11.7 Outlier region:0.8
	1R4Q_A	87.29	91.92	Favoured region:85.9 Allowed region:13.4 Outlier region:0.6
	1DM0_A	75.49	100	Favoured region:85.3 Allowed region:14.2 Outlier region:0.6

Overall, the correctness of the Vtx1-1 protein structure globally was measured by LG score of 6.716 with a MaxSub score of 0.541 to assess the quality of the protein structure while Vtx2 obtained LG score of 5.794 and a MaxSub score of 0.525. Therefore, the final selected models are Vtx1-1 (4M1U_A) and Vtx2 (4P2C_A) based on their overall stability in Verify 3D, ERRAT, and PROCHECK, and other assessment tools. These final models were successively used in further analysis.

Model refinement

The initial protein structure model may not always be accurate (Ganugapati & Akash, 2017), and this comparative model has to go for further validation for better accuracy. In this study, the Vtx structure was improved using multiple similar known templates. The structure may still contain errors due to differences (insertions and gaps) in amino acids amongst target and templates during the absences of biomolecules (proteins, nucleic acid and ligands) interactions (Feig, 2017). In comparative modelling, the primary protein structure has to refine to achieve enhanced accuracy (Heo & Feig, 2018). In the current study, the Vtx structures were refined through molecular dynamics simulation. Molecular dynamics refine the structure of the proteins and shapes a group of conformities to their native state. (Raval et al., 2012). Hence, the stability of the Vtx structure was estimated as a function of simulation time. The value of RMSD for Vtx1-1 was consistently deviated at ~2.5 Å after 5 ns till 40 ns while fluctuated between 2 Å and 2.5 Å at 40 ns to 50 ns. Consequently, Vtx2 reached a plateau state at around 2.0 Å after 10 ns and consistently deviated between 2 Å-2.5 Å until the end of the simulation. Overall, RMSD values of Vtx1-1 and Vtx2 fluctuated and deviated between 2.0-2.5 Å after 5 ns (Figure 1(a)).

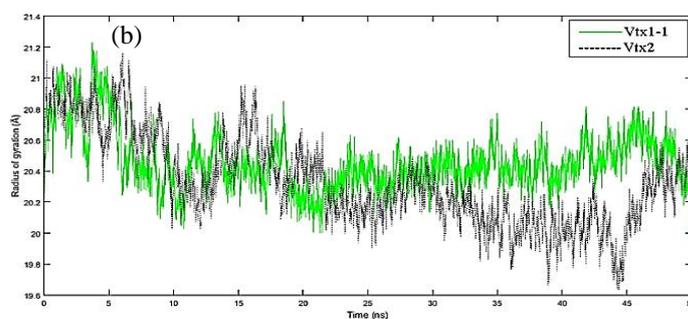
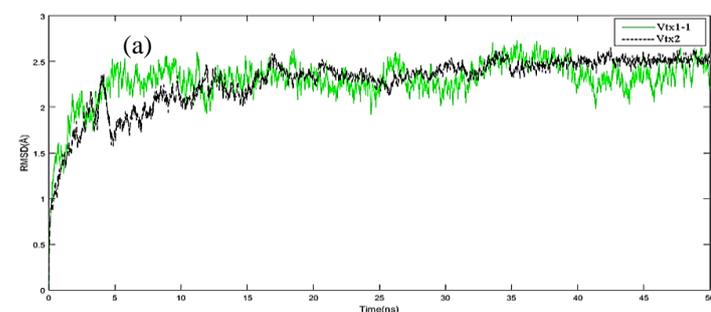


Figure 1 RMSD (a) indicate the overall stability of Vtx1-1, and Vtx2 during the simulation. Radius of gyration (b) representing the overall dimension of Vtx1-1 and Vtx2 was as a function of time

The simulation time is sufficient based on the RMSD value to facilitate the Vtx structure to reach a stabilised state. Furthermore, the radius of gyration was evaluated for the stability of the Vtx models. The compactness of a protein structure was determined through the radius of gyration (Lobanov et al., 2008). Gyration values revealed that the compactness of Vtx1-1 (20. Å-21.2 Å) and Vtx2 (19.6 Å -21.2 Å) was maintained without substantial drift, signifying the predicted Vtx structure throughout the simulation (Figure 1 (b)).

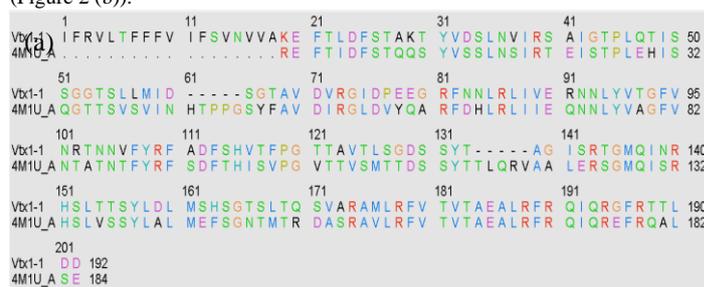
The quality of the initial Vtx models structure was refined by MODELLER prior to the *locPREFMD* refinement process to get the most suitable model for further analysis. Verify 3D showed that the percentages of amino acid with correct 3D fold (Vtx1-1: 87.97 %; Vtx2: 80.00 %). Besides that, ERRAT analysis also indicated on nonbonded interactions within the Vtx model (Vtx1-1: 91.59 %; and Vtx2: 82.75 %). The PROCHECK Ramachandran plot indicated that 100 % and 99.3 % of the Vtx1-1 and Vtx2 amino acid residues are located in preferred, approved areas, respectively.

Table 4 Evaluations between the initial models and refined models

Model	Energy (kJ/mol)	Verify 3D (%)	ERRAT (%)	PROCEHCK (%)
Vtx1-1 initial	-18752.408	93.82	97.76	Favoured regions:93.4 Allowed regions:6.6 Outlier regions:0.0
Vtx1-1 refined	-18555.111	87.97	91.59	Favoured regions:93.3 Allowed regions:6.7 Outlier regions:0.4
Vtx2 initial	15378.495	94.95	87.28	Favoured regions:87.6 Allowed regions:11.7 Outlier regions:0.8
Vtx2 refined	17386.025	80.00	82.75	Favoured regions:85.8 Allowed regions:13.5 Outlier regions:0.7

Furthermore, the energy level of Vtx1-1 and Vtx2 was -18555.111 kJ/mol and -17386.025 kJ/mol, respectively after refinement. Native protein frequently folds into conformation with the lowest energy, which is the most constant form (Kazlauskas, 2018). These indicate that refined Vtx models via MODELLER (built-in Chimera 1.14) are closer and more related to the native models (Table 4).

Vtx1-1 and Vtx2 protein were modelled with the best match selections from the PDB structure database. In this study, Vtx1-1 was modelled in accordance with 4m1u_A while Vtx2 with 4p2c_A. Alignment of the new modelled structure of Vtx1-1 with 57.07 % identity (Figure 2 (a)) while Vtx2 with 89.47 % identity (Figure 2 (b)).



(b)

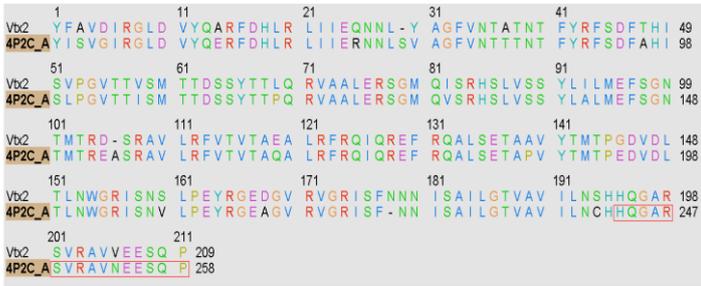


Figure 2 Alignment of new modelled structure Vtx1-1 with the native structure 4M1U_A (a) and (b) new modelled structure Vtx2 with native structure 4P2C

Alignment of both Vtx1-1 and Vtx2 was assessed by Needleman-Wunsch using BLOSUM-62 (Figure 3). Evaluation superpositions across in the final alignment overall RMSD: 0.797 and Q-score: 0.924 at cutoff 5.0 while 279 residues pairs were aligned. This alignment defines the secondary structure of proteins with the 86.87 % identity.

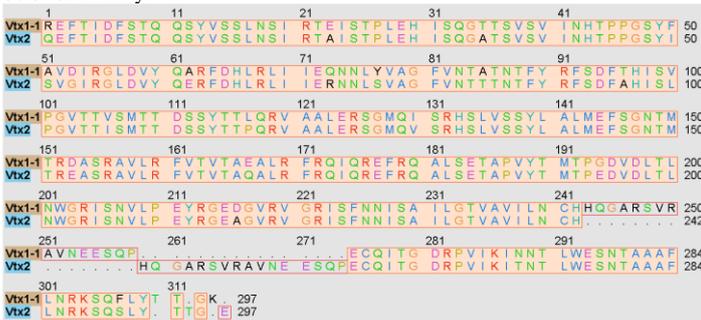


Figure 3 Alignment of the superimposed new structured model Vtx1-1 and Vtx2
Secondary structure prediction

All the modelled plasmid-mediated verotoxin gene of MN696158 (Vtx1-1) and MN688720 (Vtx2) contain 10 alpha-helices (H1-H10) and six β -strands (E1-E10) as shown in Fig 4. Despite the variations in amino acid residues in each VTX sequences (MN696158 and MN688720), the formation of secondary structures revealed a consensus between each verotoxin models (Vtx1-1 and Vtx2). This is because the protein structure is well preserved compared to the amino acid sequence throughout evolution (Dong et al., 2018). In addition, there is a preference for the collection of amino acid residues within a secondary structure (Figure 5). It can be shown that the residues of amino acids T, I, V, E and F are extremely abundant in the alpha-helix regions. These residues are classified as Helix Formats due to their low energy cost for helix formation. In the other hand, the residues of amino acids Q, S, R, T, L, V, A and Y are prevalent in beta strands due to the ability of their hydrophobic side chains to stabilize the β structure (Merkel et al., 1999).

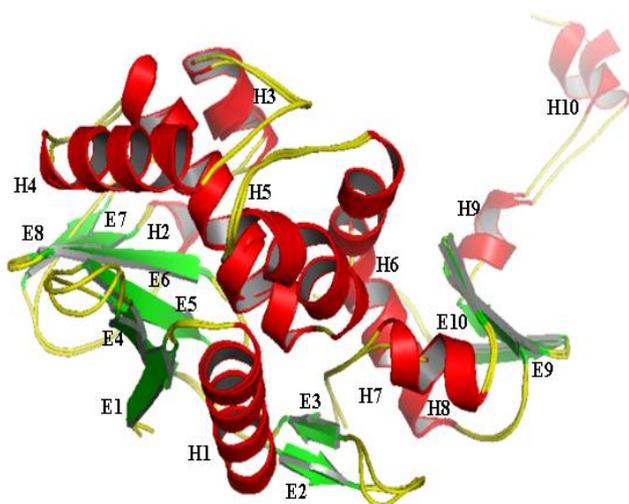
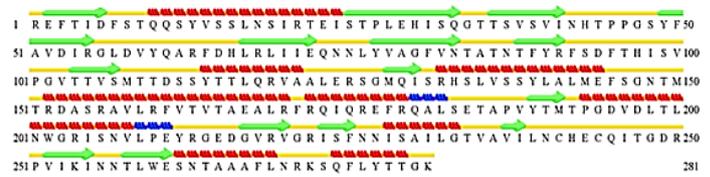


Figure 4 3D structure superimposition of Vtx1-1 and Vtx2. The α -helices (H1-H10), β -strands (E1-E10) and loops of Vtx are coloured in red, green, and yellow respectively

Legend of secondary structure icons:

- H Alpha-Helix
- E Extended Configuration (Beta-sheet)
- B Isolated Beta Bridge
- b Isolated Beta Bridge (Type 3 Fig 4,c,d)
- T Turn
- C or " " Coil
- G 3-10 Helix
- I Pi-Helix

Vtx1-1



Vtx2

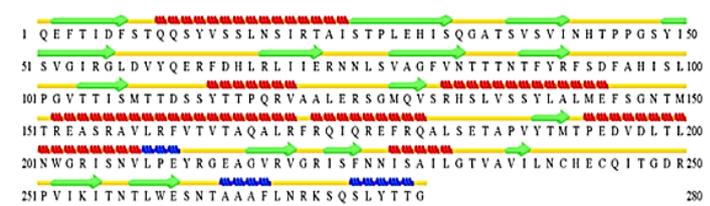


Figure 5 The legend of secondary structure (Vtx1-1 and Vtx2)

Root-mean-square-deviation

In this current analysis, RMSD values reflect VTEC's different thermal motion at the temperature concerned. Vtx1-1 The molecular dynamics simulation review (Figure 6 (a)) showed that Vtx1-1 retained an RMSD value of 3 Å to 5 ns at 303 K and fluctuated from 1.5 Å to 2.5 Å to 30 ns and remained constant at 2 Å until the end of the simulation. The RMSD value of Vtx1-1 at 313 K reached a plateau state at 2 Å at 20 ns and maintained a constant deviation between 2.5 Å and 3 Å for the last 80 ns of the simulation. Vtx1-RMSD 1's value has steadily deviated from 1 Å ~2 Å to 65 ns at 323 K. Slightly inclined to 3 Å at 65 ns and minor variations between 1 Å and 2 Å during the simulation were also observed. At 303 K, on the other hand, the RMSD value of Vtx2 (Figure 6 (b)) fluctuated marginally between 1. Å -2.5 Å to 25 ns and held at 1.5 Å until 100 ns of simulation continuously deviated. While Vtx2 showed a fluctuation between 1Å-3.5 Å and 1. at 313 K, to 25 ns and steadily deviated to ~2 Å for the rest of the simulation process.

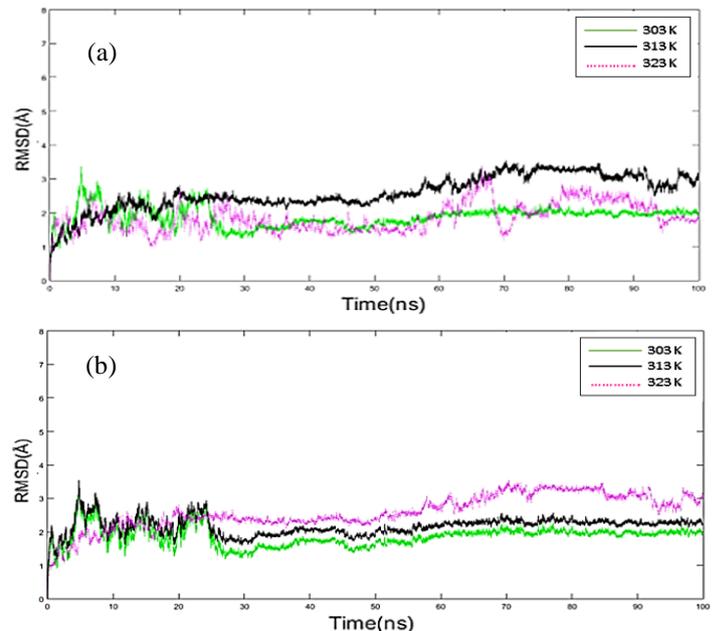


Figure 6 Root-mean-square deviation (RMSD) of the protein backbone as a time function at 303 K, 313 K and 323 K of (a)Vtx1-1 and (b) Vtx2

Furthermore, after 5 ns, the RMSD value of Vtx2 entered a plateau state and continuously deviated from 2.5 Å -3 Å until 100 ns of simulation at 323 K. During MD simulation, RMSD modifications suggest that the protein backbone of all VTEC structures has moved to a new conformation to preserve stability and versatility at different temperatures (Fields et al., 2015). The overall changes in RMSD (1.5-2.0 Å) are considered minimal because the Vtx protein

conformations (with the exception of Vtx2 at 323 K) are depicted closely in line with their initial structures (Kato et al., 2017). The RMSD outcome therefore shows that the temperature has the least effect on Vtx1-1 and Vtx2, respectively. In addition, no significant deviation was observed at the time of the MD simulation of 100 ns after the RMSD plateau was reached by VTEC to allow VTEC to obtain a new constant conformation.

Radius of gyrations

Protein compactness is another way to calculate a protein's stability (Paul et al., 2014). From a Rg study, the impact of temperature on the VTEC dimension was assembled. At all temperatures, Vtx1-1 displayed a similar dimension of ~19 Å–20 Å. Meanwhile, the Rg value of Vtx1-1 was stretched between ~20 Å until 100 ns at 303 K, which corresponds to its RMSD changes that occurred at the same time frame. The Rg value was also held at 313 K and 323 K respectively at a steady value of 19.5–20.6 Å (Figure 7 (a)). At all temperatures, the Rg value of Vtx2 has decreased to 19.5 Å from 21.5 Å between 20–100 ns (Figure 7 (b)). Overall, during the simulation, the Rg of VTEC models remained stable, indicating that VTEC structures are successful in preserving their original compactness even when the temperature rises.

Consequently, during the MD simulation study, the entire size of Vtx1-1 remained constant, although exceptions were observed on Vtx2. The shifts in Rg suggest that temperature plays a crucial role in the loosening of the protein system's molecular structural network and/or collapse at 323 K (Gu et al., 2019). The VTEC compactness analysis via Rg, similar to the RMSD analysis, also indicates that the Vtx1-1 model presents maximum stability at higher temperatures than Vtx2.

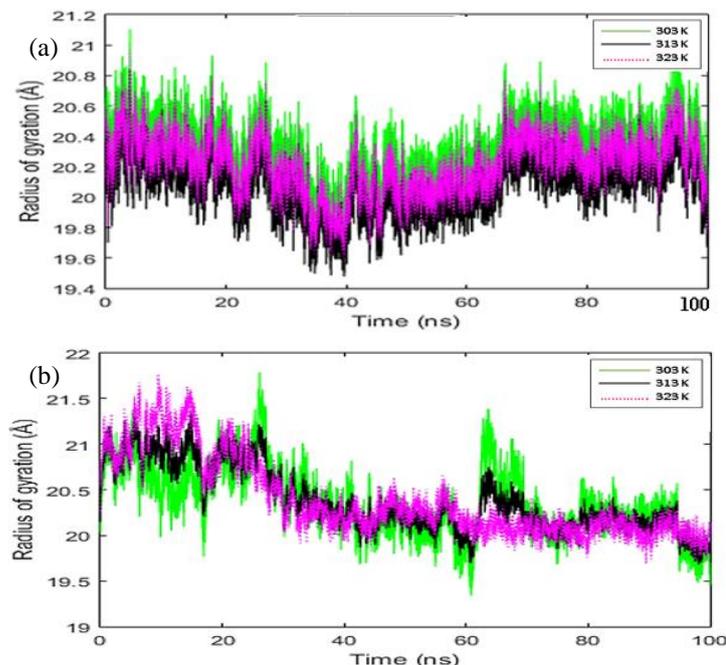


Figure 7 Radius of the gyration as a time function at 303 K, 313 K and 323 K of (a) Vtx1-1 and (b) Vtx2

A recent study shows that the hydrophobic effect and temperature play a major role in protein folding (Pucci & Rooman, 2017). Hydrophilic residues are often present under normal conditions on the protein surface, while hydrophobic residues are often suppressed away from the aqueous environment within the protein (Ramli et al., 2018). The hydrophobic region is exposed to the solvent during the protein denaturation process (the enzyme loses its catalytic activity) (Paul et al., 2014). Soluble proteins, in interaction with the solvent, minimise the surface (Malleshappa et al., 2014).

Solvent accessible surface area

The hydrophobic effect is driven by protein folding of the solvent accessible surface area and is temperature dependent (Pucci & Rooman, 2017). Under normal conditions, hydrophilic residues are usually located on the protein surface, while hydrophobic residues are generally buried within the protein away from the aqueous environment (Ramli et al., 2018). The solvent would be exposed to the hydrophobic region (Paul et al., 2014) if protein denaturation were to occur. Changes in SASA investigated the influence of temperature on the distribution of hydrophilic and hydrophobic residues of VTX models. In Vtx1-, SASA's value ranged from ~106.10–114.81 nm² at 303 K, ~107.05–113.10 nm² at 313 K and ~102.23–112.18 nm² at 323 K (Figure 8 (a)). Meanwhile, the Vtx2 SASA retained at 303 K between ~117.03–125.10 nm², at 313 K between ~113.63–126.02 nm². Vtx2 maintained a stable SASA at 323K between ~115.60–125.31 nm² and maintained an average range of 108.12 nm² until the simulation

was completed (Figure 8 (b)).

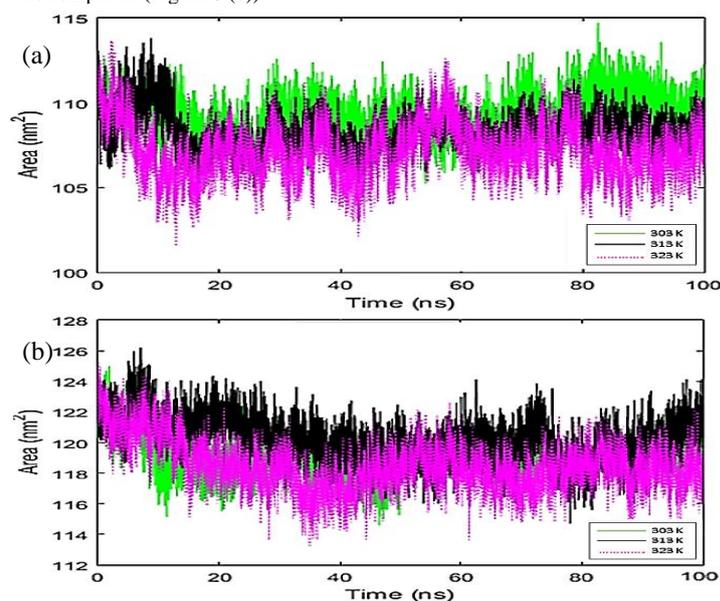


Figure 8 Solvent accessible surface area (SASA) of (a) Vtx1-1 and (b) Vtx2 as a time function at 303 K, 313 K and 323 K

The SASA study of VTEC models showed a pattern close to that presented during the simulation phase by RMSD and Rg. The SASA study of Vtx1-1 and Vtx2 showed stability at 303 K and 313 K respectively (native fold without disruption). The SASA study of Vtx1-1 and Vtx2 was, by comparison, expanded, killed or unfolded at 323 K. This enlargement of Rg showed that within the hydrophobic heart, hydrophobic molecules were distributed on the surface instead of suppressed. Furthermore, the decrease in SASA analysis is associated with the overall decrease in scale, resulting in high-temperature accumulation (Rosa et al., 2017). Residues congregate within the solvent protein, resulting in a decrease in SASA during protein aggregation (Mishra et al., 2018).

Intramolecular hydrogen bonds

The stability of the preservation of the protein structure is based on temperature-dependent interactions of the hydrogen bond (Pace et al., 2014). The hydrophobic effects, on the other hand, are the folding of the overall protein structure, since the hydrogen bond is directly connected to the protein structure (secondary and tertiary structure) and protein interaction selectivity (Gao et al., 2015). The Vtx1-1 and Vtx2 starting structures have intramolecular hydrogen bonds of 298 and 236, respectively (Figure 9). In comparison, the lowest number of intramolecular hydrogen bonds found in the Vtx2 models refers to distortion at high temperatures in RMSD, Rg and SASA.

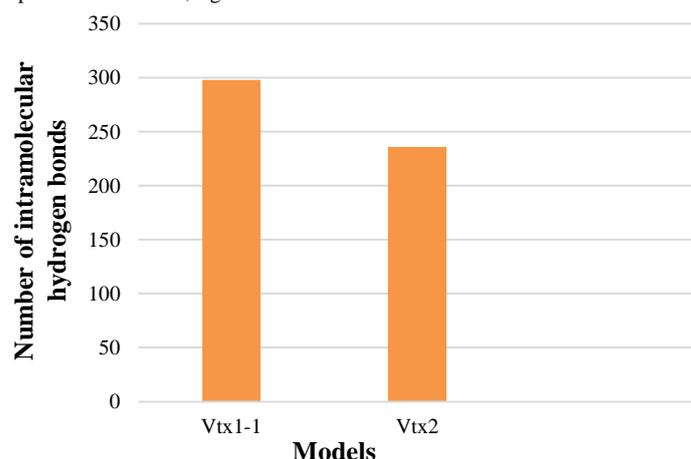


Figure 9 Number of intramolecular hydrogen bonds in the initial structure of Vtx1-1 and Vtx2

The simulation regardless of temperature plays an essential role in intramolecular hydrogen bonds. At 303 K, Vtx1-1 retained an average number of 120.97 intramolecular hydrogen bonds (Figure 10 (a)). This average number further decreased to 113.94 at 313 K and increased to 124.26 at 323 K respectively. In addition, Vtx2 has an average number of 137.37 intramolecular hydrogen bonds at 303 K, 138.30 at 313 K and 142.30 at 323 K, respectively (Figure 10 (b)).

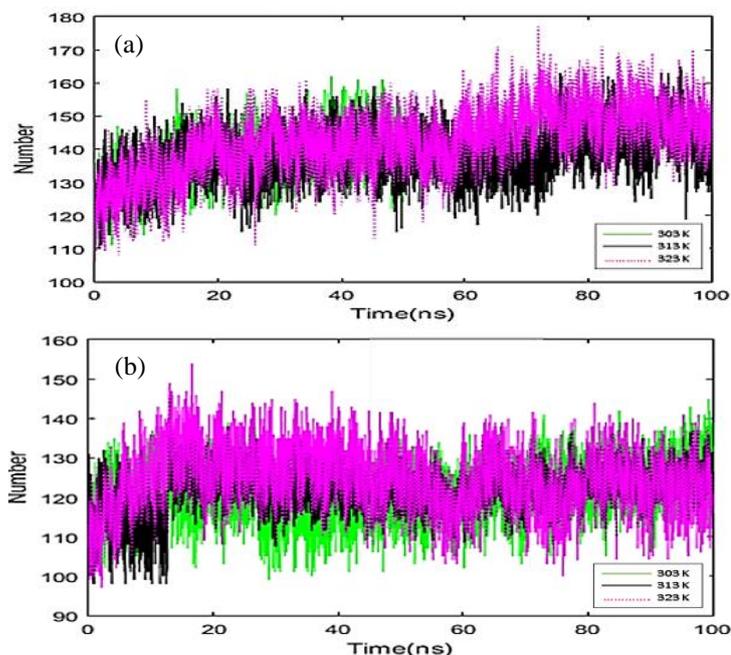


Figure 10 Number of hydrogen intramolecular bonds of (a) Vtx1-1 and (b) Vtx2 as a function of time at 303 K, 313 K and 323 K

This result specifies the hydrogen bond network within the Vtx structures. Besides, the Vtx structure stability was disrupted and weakens as the temperature rises in accordance with a recent study (Mallamace et al., 2018).

Root-mean-square fluctuation

At residue level, the stability of verotoxin (VTX) was also experienced. The flexibility and mobility of protein residues at various temperatures has been demonstrated by RMSF. In Vtx1-1, the fluctuation was recorded at residue position of 4-8, 45-48, 121-124, and 190-192 at 303 K (Figure 11 (a)). These residues also fluctuated at 4-12, 43-52 at 313K and 44-51 at 323K, respectively. In the meantime, Vtx2 demonstrated versatility at different residue positions; 198-207 at 303 K, 5-12, 21-26, 40-47, 197-199, 313 K at 200-205 and 323 K at 9-11, 157-160, 161-165, 204-208 (Figure 11 (b)).

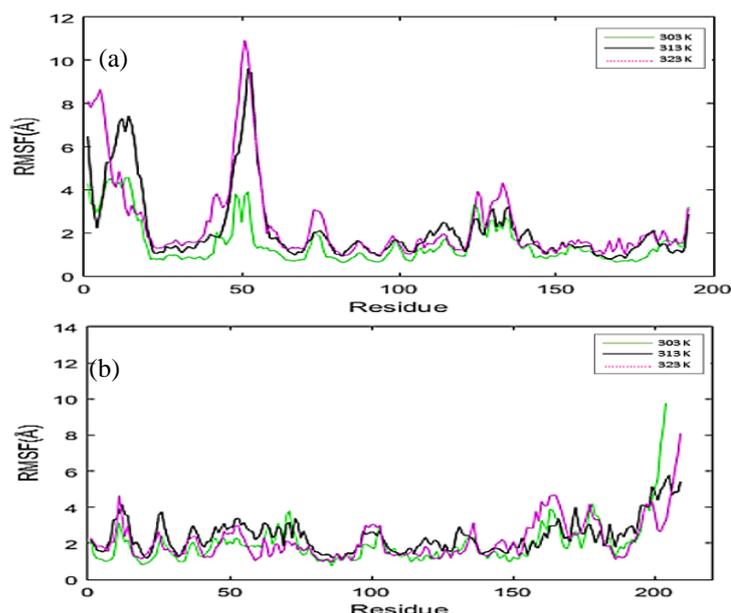


Figure 11 RMSF of (a) Vtx1-1 and (b) Vtx2 residues number at 303 K, 313 K and 323 K.

After the first 100 protein residues in the Vtx2 model, residue variations were more prominent. At the same time, the residues for the remaining position were relatively less fluctuating. Higher RMSF values suggest greater flexibility in conformation and may become unstable as the temperature increases (Du et al., 2017).

The physicochemical properties of the residues of amino acids affect protein stability and versatility (Alvarez-Ponce et al., 2018). From the thermal sensitive regions, the effect of amino acid composition on Vtx1-1 sequence information was obtained. TIDFSTQQS, HTPPGSYFAV, AALE, TMT, IDFSTQQS,

RTAIST, VINHTPPG, AVLRFVTVT, and DLTLNWGRISNV are the temperature-sensitive sequences listed (Table 5). Fluctuations in the MD simulation study is responsible for the most common amino acid residues: asparagine, aspartic acid, cysteine, lysine, glutamine, glycine, serine, threonine and tyrosine. Asparagine, cysteine, glutamine, glycine, serine, threonine and tyrosine are non-charged among the amino acid residues known and are unable to define sufficient hydrogen bonds to stabilise structures (Ramli et al., 2012).

Table 5 Sequence ranges on the fluctuated zones

Model	Fluctuated region	Sequence
Vtx1-1	4-12	TIDFSTQQS
	43-52	HTPPGSYFAV
	121-124	AALE
	190-192	TMT
Vtx2	5-12	IDFSTQQS
	21-26	RTAIST
	40-47	VINHTPPG
	157-165	AVLRFVTVT
	197-208	DLTLNWGRISNV

In summary, MN696158 (Vtx1-1) and MN688720 (Vtx2) showed key differences in their respective versatility of conformation and stability of temperature rise. MN696158 (Vtx1-1) is the most thermostable VTX model exhibited during the simulation process due to its least deviated dynamics. In the meantime, RMSD shows that as temperature rises, MN688720 (Vtx2) is the least stable model of plasmid-mediated VTX models. On the other hand, in Rg, SASA and differences between intramolecular and intermolecular hydrogen bonds, Vtx2 showed denature activity. In contrast, RMSF analysis exhibited that the initial fluctuated regions are protein loops located in the L-subdomain with a temperature increase due to lack of hydrophobicity and electrostatic interfaces. The results of the current study provide a deeper understanding of protein structural analysis in an essential region of molecular synthesis and will greatly improve successful and focused intervention approaches.

CONCLUSION

In this study, we reported structural analysis of two plasmid-mediated verotoxin gene sequences. Despite these plasmid-mediated verotoxin gene sequences showed slight differences in sequence level. The 3D verotoxin models produced demonstrated remarkable variations in their relative versatility of compliance and stability when the temperature increased. Protein loops on the L-subdomain are among the key fluctuated temperature-increasing regions because of the lack of electrostatic interactions and hydrophobicity in this region. It is predicted that a thermal, stable plasmid mediated verotoxin gene model can be best suitable model for the predictions of vaccine developments through the protein engineering process of structural information obtained.

Conflict of study: None

Acknowledgement: We are grateful to Universiti Malaysia Pahang and Hospital Tengku Ampuan Afzan, Malaysia for research facilities. We also thank Medical Research & Ethics Committee (MREC), Malaysia for ethical approval. Author [Md. Fazlul Karim Khan] is thankful to Universiti Malaysia Pahang for providing Doctoral Research Scholarship (DRS) and Postgraduate Research Grants Scheme (PGRS 180360) as a financial support.

REFERENCES

Amir, A., Kapoor, N., Kumar, H., Tariq, M., & Siddiqui, M. A. (2019). In silico homology modeling and epitope prediction of drug target protein in human herpes virus 8 (HHV8). *Biotech Today: An International Journal of Biological Sciences*, 9(1), 41-48. <http://dx.doi.org/10.5958/2322-0996.2019.00005.X>

Chakkyarath, V., & Natarajan, J. (2019). Identification of Ideal Multi-targeting Bioactive Compounds Against Mur Ligases of Enterobacter aerogenes and Its Binding Mechanism in Comparison with Chemical Inhibitors. *Interdisciplinary Sciences: Computational Life Sciences*, 11(1), 135-144. <http://dx.doi.org/10.1007/s12539-017-0261-4>

Chakravarty, S., Godbole, S., Zhang, B., Berger, S., & Sanchez, R. (2008). Systematic analysis of the effect of multiple templates on the accuracy of comparative models of protein structure. *BMC structural biology*, 8(1), 31

Cheung, V., & Trachtman, H. (2014). Hemolytic uremic syndrome: toxins, vessels, and inflammation. *Front Med (Lausanne)*, 1, 42. <http://dx.doi.org/10.3389/fmed.2014.00042>

Cloete, R., Kapp, E., Joubert, J., Christoffels, A., & Malan, S. F. (2018). Molecular modelling and simulation studies of the Mycobacterium tuberculosis multidrug efflux pump protein Rv1258c. *PloS One*, 13(11), e0207605. <http://dx.doi.org/10.1371/journal.pone.0207605>

Colello, R., Caceres, M. E., Ruiz, M. J., Sanz, M., Etcheverria, A. I., & Padola, N. L. (2016). From farm to table: follow-up of Shiga toxin-producing

- Escherichia coli* throughout the pork production Chain in Argentina. *Frontiers in Microbiology*, 7, 93. <http://dx.doi.org/10.3389/fmicb.2016.00093>
- Dong, R., Pan, S., Peng, Z., Zhang, Y., & Yang, J. (2018). mTM-align: a server for fast protein structure database search and multiple protein structure alignment. *Nucleic acids research*, 46(W1), W380-W386
- Dorn, M., e Silva, M. B., Buriol, L. S., & Lamb, L. C. (2014). Three-dimensional protein structure prediction: Methods and computational strategies. *Computational Biology and Chemistry*, 53, 251-276. <http://dx.doi.org/10.1016/j.compbiolchem.2014.10.001>
- Duanis-Assaf, D., Kenan, E., Sionov, R., Steinberg, D., & Shemesh, M. (2020). Proteolytic Activity of *Bacillus subtilis* upon κ -Casein Undermines Its "Caries-Safe" Effect. *Microorganisms*, 8(2), 221
- Elengoe, A., Naser, M. A., & Hamdan, S. (2014). Modeling and docking studies on novel mutants (K71L and T204V) of the ATPase domain of human heat shock 70 kDa protein 1. *International Journal of Molecular Sciences*, 15(4), 6797-6814. <http://dx.doi.org/10.3390/ijms15046797>
- Faure, G., Joseph, A. P., Craveur, P., Narwani, T. J., Srinivasan, N., Gelly, J. C., Rebehmed, J., & de Brevern, A. G. (2019). iPBAvizu: a PyMOL plugin for an efficient 3D protein structure superimposition approach. *Source Code for Biology and Medicine*, 14(1), 5. <http://dx.doi.org/10.1186/s13029-019-0075-3>
- Fazlul, M., Rashid, S. S., Nazmul, M., Zaidul, I., Baharudin, R., & Nor, A. (2018). A clinical update on antibiotic resistance gram-negative bacteria in Malaysia-A review. *Journal of International Pharmaceutical Research*, 45, 270-283
- Fazlul, M., Zaini, M., Rashid, M., & Nazmul, M. (2011). Antibiotic susceptibility profiles of clinical isolates of *Pseudomonas aeruginosa* from Selayang Hospital, Malaysia. *Biomedical Research*, 22(3), 263-266. <https://doi.org/10.1016/j.ijid.2012.05.589>
- Feig, M. (2017). Computational protein structure refinement: Almost there, yet still so far to go. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 7(3), e1307. <http://dx.doi.org/ARTN E1307 10.1002/wcms.1307>
- Frith, M. C. (2019). How sequence alignment scores correspond to probability models. *Bioinformatics*, 36(2), 408-415. <http://dx.doi.org/10.1093/bioinformatics/btz576>
- Ganugapati, J., & Akash, S. (2017). Multi-template homology based structure prediction and molecular docking studies of protein 'L' of Zaire ebolavirus (EBOV). *Informatics in Medicine Unlocked*, 9, 68-75
- Geng, H., Chen, F., Ye, J., & Jiang, F. (2019). Applications of molecular dynamics simulation in structure prediction of peptides and proteins. *Computational and Structural Biotechnology Journal*, 17, 1162-1170. <http://dx.doi.org/10.1016/j.csbj.2019.07.010>
- Haddad, Y., Adam, V., & Heger, Z. (2020). Ten quick tips for homology modeling of high-resolution protein 3D structures. *PLoS Computational Biology*, 16(4), e1007449. <http://dx.doi.org/10.1371/journal.pcbi.1007449>
- Hai-You, D., Ya, J., & Yang, Z. (2016). Protein structure prediction. *Acta Physica Sinica*, 65(17)
- Hazards, E. P. (2013). Scientific opinion on VTEC-seropathotype and scientific criteria regarding pathogenicity assessment. *EFSA Journal*, 11(4), 3138
- Heo, L., & Feig, M. (2018). Experimental accuracy in protein structure refinement via molecular dynamics simulations. *Proceedings of the National Academy of Sciences*, 115(52), 13276-13281. <http://dx.doi.org/10.1073/pnas.1811364115>
- Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B. L., Grubmüller, H., & MacKerell, A. D. (2017). CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature methods*, 14(1), 71-73
- Jing, X., & Dong, Q. (2017). MQAPRank: improved global protein model quality assessment by learning-to-rank. *BMC Bioinformatics*, 18(1), 275. <http://dx.doi.org/10.1186/s12859-017-1691-z>
- Kaper, J. B., & O'Brien, A. D. (2014). Overview and historical perspectives. *Microbiology Spectrum*, 2(6). <http://dx.doi.org/10.1128/microbiolspec.EHEC-0028-2014>
- Kazlauskas, R. (2018). Engineering more stable proteins. *Chemical Society Reviews*, 47(24), 9026-9045. <http://dx.doi.org/10.1039/C8CS00014J>
- Kerfeld, C. A., & Scott, K. M. (2011). Using BLAST to teach "E-value-tionary" concepts. *PLoS Biology*, 9(2), e1001014. <http://dx.doi.org/10.1371/journal.pbio.1001014>
- Khan, M. F. K., Kabir, M. N., Rashid, S. S., Tayan, O., & Maziz, M. N. H. (2020). *Computational Investigation on Protein Sequence of Non-O157 VTEC for Potentiality of Vaccine Production*. Paper presented at the 2020 IEEE International Conference on Automatic Control and Intelligent Systems (ICACIS).
- Khan, M. F. K., Rashid, S. S., Maziz, M. N. H., Ramli, A. N. M., & Baharuddin, R. (2020). Plasmid-mediated verotoxin producing non-O157 *Escherichia coli* isolates from Malaysia. *International Journal of Infectious Diseases*, 101, 138. <http://dx.doi.org/10.1016/j.ijid.2020.09.374>
- Khan, M. F. K., Rashid, S. S., Ramli, A. N. M., Maziz, M. N. H., & Roesnita, B. (2020). Molecular Characterization of Plasmid-Mediated Non-O157 Verotoxigenic *Escherichia coli* Isolated from Infants and Children with Diarrhea. *Baghdad Science Journal*, 17(3), 0710-0710. <http://dx.doi.org/10.21123/bsj.2020.17.3.0710>
- Lobanov, M. Y., Bogatyreva, N., & Galzitskaya, O. (2008). Radius of gyration as an indicator of protein structure compactness. *Molecular Biology*, 42(4), 623-628. <http://dx.doi.org/10.1134/S0026893308040195>
- Mallamace, D., Fazio, E., Mallamace, F., & Corsaro, C. (2018). The role of hydrogen bonding in the folding/unfolding process of hydrated lysozyme: A review of recent NMR and FTIR results. *International Journal of Molecular Sciences*, 19(12), 3825. <http://dx.doi.org/10.3390/ijms19123825>
- Mei, G. Y., Tang, J., Carey, C., Bach, S., & Kostrzynska, M. (2015). The effect of oxidative stress on gene expression of Shiga toxin-producing *Escherichia coli* (STEC) O157:H7 and non-O157 serotypes. *International Journal of Food Microbiology*, 215, 7-15. <http://dx.doi.org/10.1016/j.ijfoodmicro.2015.07.029>
- Merkel, J. S., Sturtevant, J. M., & Regan, L. (1999). Sidechain interactions in parallel β sheets: the energetics of cross-strand pairings. *Structure*, 7(11), 1333-1343
- Monzon, A. M., Zea, D. J., Marino-Buslje, C., & Parisi, G. (2017). Homology Modeling in a Dynamical World. *Protein Science*, 26(11), 2195-2206. <http://dx.doi.org/10.1002/pro.3274>
- Muhammed, M. T., & Aki-Yalcin, E. (2019). Homology modeling in drug discovery: Overview, current applications, and future perspectives. *Chemical Biology & Drug Design*, 93(1), 12-20. <http://dx.doi.org/10.1111/cbdd.13388>
- Nazmul, M., Fazlul, M., & Rashid, M. (2012). Plasmid profile analysis of non-O157 diarrheagenic *Escherichia coli* in Malaysia. *Indian Journal of Science*, 1(2), 130-132
- Paul, M., Hazra, M., Barman, A., & Hazra, S. (2014). Comparative molecular dynamics simulation studies for determining factors contributing to the thermostability of chemotaxis protein "CheY". *Journal of Biomolecular Structure and Dynamics*, 32(6), 928-949
- Public Health England. (2018). Epidemiology of Non-O157 Shiga toxin producing *Escherichia coli* (STEC). Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/732577/non-O157_STEC_Evidence_Base.pdf
- Pucci, F., & Rooman, M. (2017). Physical and molecular bases of protein thermal stability and cold adaptation. *Current Opinion in Structural Biology*, 42, 117-128. <http://dx.doi.org/10.1016/j.sbi.2016.12.007>
- Ramli, A. N. M., Manas, N. H. A., Hamid, A. A. A., Hamid, H. A., & Ilias, R. M. (2018). Comparative structural analysis of fruit and stem bromelain from *Ananas comosus*. *Food Chemistry*, 266, 183-191. <http://dx.doi.org/10.1016/j.foodchem.2018.05.125>
- Raval, A., Piana, S., Eastwood, M. P., Dror, R. O., & Shaw, D. E. (2012). Refinement of protein structure homology models via long, all-atom molecular dynamics simulations. *Proteins: Structure, Function, and Bioinformatics*, 80(8), 2071-2079. <http://dx.doi.org/10.1002/prot.24098>
- Sajib, S. A., Billah, M., Mahmud, S., Miah, M., Hossain, F., Omar, F. B., Roy, N. C., Hoque, K. M. F., Talukder, M. R., & Kabir, A. H. (2020). Plasma activated water: the next generation eco-friendly stimulant for enhancing plant seed germination, vigor and increased enzyme activity, a study on black gram (*Vigna mungo* L.). *Plasma Chemistry and Plasma Processing*, 40(1), 119-143. <http://dx.doi.org/10.1007/s11090-019-10028-3>
- Sefidbakht, Y., Ranaei Siadat, O., & Taheri, F. (2017). Homology modeling and molecular dynamics study on *Schwanniomyces occidentalis* alpha-amylase. *Journal of Biomolecular Structure and Dynamics*, 35(3), 574-584. <http://dx.doi.org/10.1080/07391102.2016.1154892>
- Soumiya, S., Santhiagu, A., Manjusha, C., Adarsh, V., & Prakash, S. J. (2019). Homology modeling and structural validation of gel c gene involved in the biosynthesis of gellan from *Sphingomonas paucimobilis* ATCC 31461. *Research Journal of Pharmacy and Technology*, 12(3), 1044-1050
- Su, X. D., Zhang, H., Terwilliger, T. C., Liljas, A., Xiao, J., & Dong, Y. (2015). Protein crystallography from the perspective of technology developments. *Crystallography Reviews*, 21(1-2), 122-153. <http://dx.doi.org/10.1080/0889311X.2014.973868>
- Tran, N. T., Jakovlić, I., & Wang, W.-M. (2015). In silico characterisation, homology modelling and structure-based functional annotation of blunt snout bream (*Megalobrama amblycephala*) Hsp70 and Hsc70 proteins. *Journal of Animal Science and Technology*, 57(1), 44. <http://dx.doi.org/10.1186/s40781-015-0077-x>
- Webb, B., & Sali, A. (2017). Protein structure modeling with MODELLER. In *Functional Genomics* (pp. 39-54): Springer.
- Wei, D. Q., Ma, Y., Cho, W. C. S., Xu, Q., & Zhou, F. (2017). *Translational bioinformatics and its application*: Springer Netherlands.
- Zhu, F., Zhou, Y. K., Ji, Z. L., & Chen, X. R. (2018). The Plant ribosome-inactivating proteins play important roles in defense against pathogens and insect pest attacks. *Frontiers in Plant Science*, 9, 146. <http://dx.doi.org/10.3389/fpls.2018.00146>