

Immunoinformatics assisted design of a multi-epitope kit for detecting *Cronobacter sakazakii* in powdered infant formula

Oladipo, Elijah K; Obisanya, Olukayode I; Owoeye, Victoria O; Shittu, Oyinlola G; Adamitonde, Mautin G; Ikwuka, Emmanuel C; Ojewale, Solomon O; Tijani, Adewale J; Adedokun, Feranmi A; Adedokun, Amudatu A; Oyedepo, Temitope A; Onyeaka, Helen

DOI:

[10.1093/fqsafe/fyae005](https://doi.org/10.1093/fqsafe/fyae005)

License:

Creative Commons: Attribution-NonCommercial (CC BY-NC)

Document Version

Publisher's PDF, also known as Version of record

Citation for published version (Harvard):

Oladipo, EK, Obisanya, OI, Owoeye, VO, Shittu, OG, Adamitonde, MG, Ikwuka, EC, Ojewale, SO, Tijani, AJ, Adedokun, FA, Adedokun, AA, Oyedepo, TA & Onyeaka, H 2024, 'Immunoinformatics assisted design of a multi-epitope kit for detecting *Cronobacter sakazakii* in powdered infant formula', *Food Quality and Safety*, vol. 8, fyae005. <https://doi.org/10.1093/fqsafe/fyae005>

[Link to publication on Research at Birmingham portal](#)

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Article

Immunoinformatics assisted design of a multi-epitope kit for detecting *Cronobacter sakazakii* in powdered infant formula

Elijah K. Oladipo^{1,2,3}, Olukayode I. Obisanya^{2,4}, Victoria O. Owwoye^{2,5}, Oyinlola G. Shittu⁶, Mautin G. Adamitonde⁷, Emmanuel C. Ikwuka⁸, Solomon O Ojewale⁸, Adewale J. Tijani⁹, Feranmi A. Adedokun⁹, Amudatu A. Adedokun¹⁰, Temitope A. Oyedepo¹¹, and Helen Onyeaka^{12,*}

¹Department of Microbiology, Laboratory of Molecular Biology, Immunology and Bioinformatics, Adeleke University, Ede, Osun State, Nigeria

²Genomics Unit, Helix Biogen Institute, Ogbomoso, Oyo State, Nigeria

³Department of Microbiology, Ladoko Akintola University of Technology, Ogbomoso, Oyo State, Nigeria

⁴Department of Medical Laboratory Science, College of Medicine, University of Lagos

⁵Department of Microbiology, Obafemi Awolowo University, Ile-Ife, Osun state, Nigeria

⁶Department of Physiology, Ladoko Akintola University of Technology, Ogbomoso, Oyo State, Nigeria

⁷Department of Human Anatomy, Ladoko Akintola University of Technology, Ogbomoso, Oyo State, Nigeria

⁸Department of Biotechnology, Federal University of Technology, Akure, Ondo State, Nigeria

⁹Department of Biochemistry, Ladoko Akintola University of Technology, Ogbomoso, Oyo State, Nigeria

¹⁰Medical Laboratory Science Department, Fountain University, Osogbo, Osun State, Nigeria

¹¹Department of Biochemistry, Adeleke University, Ede, Osun State, Nigeria

¹²School of Chemical Engineering, University of Birmingham, Edgbaston, Birmingham, UK

*Correspondence to: Helen Onyeaka, School of Chemical Engineering, University of Birmingham, Edgbaston, Birmingham B12 2TT, UK. E-mail: h.onyeaka@bham.ac.uk

Abstract

Objectives: *Cronobacter sakazakii*, formerly *Enterobacter sakazakii*, is an emerging ubiquitous and opportunistic foodborne pathogen with a high mortality rate. It has been implicated in cases of meningitis, septicaemia, and necrotizing enterocolitis among infants worldwide in association with powdered infant formula (PIF). This study was an insilico designed peptide base kit framework, using immunoinformatic techniques for quick detection of *C. sakazakii* in PIF.

Materials and Methods: In the present study, a peptide-based kit was designed with a bioinformatic technique to rapidly identify *C. sakazakii* in PIF using *flhE*, *secY*, and *bcsC*, which are genes responsible for its biofilm formation, as target genes. The antigenicity, membrane topology, and the presence of signal peptides of the target genes were analysed using VaxiJen, DeepTMHMM, and SignalP servers. To provide stability and flexibility to the multiple-epitope construct, the linear B cells and helper T cells (*IL-4* (interleukin 4) and *IL-10* (interleukin 10) inducing epitopes) were linked with a GSGSG linker followed by the addition of protein disulphide bonds. To ascertain specificity, the multi-epitope construct was molecularly docked against genes from sources other than PIF, like alfalfa, and the environment, with PIF being the highest: -328.48. Finally, the codons were modified using the pET28a(+) vector, and the resultant multi-epitope construct was successfully cloned *in silico*.

Results: The final construct had a length of 486 bp, an instability index of 23.26, a theoretical pI of 9.34, a molecular weight of 16.5 kDa, and a Z-score of -3.41.

Conclusions: The multi-epitope peptide construct could be a conceptual framework for creating a *C. sakazakii* peptide-based detection kit, which has the potential to provide fast and efficient detection. However, there is a need for additional validation through the *in vitro* and *in vivo* techniques.

Keywords: Powdered infant formula; *Cronobacter sakazakii*; bioinformatics; immunoinformatics; detection kit; peptide-based; epitopes; analytical kits; recombinant fusion protein; peptides.

Introduction

An emerging pathogen identified in powdered infant formula (PIF) consumption and linked epidemiologically to neonatal foodborne outbreaks is *Cronobacter sakazakii* (Cai *et al.*, 2013; Fei *et al.*, 2015). It is a Gram-negative, oxidase-negative, catalase-positive, and facultatively anaerobic

bacteria belonging to the Enterobacteriaceae family. They are generally motile, reduce nitrate, use citrate, and hydrolyse esculin and arginine. *Cronobacter sakazakii*, formerly *Enterobacter sakazakii*, comprises seven species: *C. sakazakii*, *C. malonaticus*, *C. turicensis*, *C. muytjensii*, *C. condimenti*, *C. universalis*, and *C. dublinensis* (Iversen *et al.*, 2007). Several

Received 3 July 2023; Revised 19 October 2023; Editorial decision 6 November 2023

© The Author(s) 2024. Published by Oxford University Press on behalf of Zhejiang University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

strains of *C. sakazakii*, *C. malonaticus*, *C. turicensis*, and *C. dublinensis* are desiccation-resistant and persistent in dried products such as PIF (Fei et al., 2022).

In addition to being ubiquitous, this organism has been isolated from a variety of sources, including hospitals, PIF production facilities, and homes, as well as from various food categories like dried food products and spices, and it is expressed in breast milk (McMullan et al., 2018). Most *C. sakazakii* infections have been reported to affect newborns, children, and people with weakened immune systems (Healy et al., 2010). Numerous countries have reported nosocomial outbreaks in neonatal intensive care units (NICUs; Caubilla-Barron et al., 2007; Hariri et al., 2013) with high mortality rates ranging from 40% to 80%.

The current methods used to detect and identify *C. sakazakii* include culture and biochemical-based methods, polymerase chain reaction or gene probe assays. These methods require highly skilled personnel and expensive equipment or longer turnaround time; hence, they could be less effective for real-time surveillance. Therefore, designing a cheaper, easy-to-use, highly sensitive, specific detection kit will be beneficial. There is also a need for more consistency between results obtained with different detection and identification methods; the current United States Food and Drug Administration detection method uses a series of culturing steps to isolate *C. sakazakii* from food matrices. Several target genes such as *flhE*, *bcsc*, and *secY* have been incriminated to be responsible for the organism's pathogenicity (Hartmann et al., 2010; Li et al., 2020). With the increasing availability of genome sequences in the NCBI database combined with bioinformatics tools, retrieving selected target gene sequences with practical verification is possible. Several studies have used this *in silico* approach to design a detection kit for specific targets for pathogen detection (Schoch et al., 2020; Oladipo et al., 2023b). However, no such study has been done for *Cronobacter* species. In this study, bioinformatics tools were extensively utilized for the *in silico* design of a detection kit to effectively detect *C. sakazakii* in PIF and other related food categories.

Methods

Sequence retrieval and antigenicity prediction

Protein Knowledgebase (UniProtKB) and the National Center for Biotechnology Information (NCBI) were adopted to identify the amino acid sequences of the targeted genes. The genes' accession number, amino acid sequence, and tertiary structure (3D) were retrieved from the database (Oladipo et al., 2023a). Using the VaxiJen 2.0 Bacteria model, it was predicted that these genes would trigger a response in the immune system (Doytchinova and Flower, 2007). VaxiJen generates protein antigenicity without using alignment and emphasizes the physical and chemical features of the identified alternative (Oladipo et al., 2023a) (Figure 1).

Signal peptide and transmembrane topology

DeepTMHMM Server (Hallgren et al., 2022) was used to predict the transmembrane topology of the selected antigenic genes and classified them into four regions: Outside, Inside, Membrane or Transmembrane, and Signal. Genes with signal peptides were further confirmed with the SignalP server (Almagro Armenteros et al., 2019). The amino acid sequence that fell within these

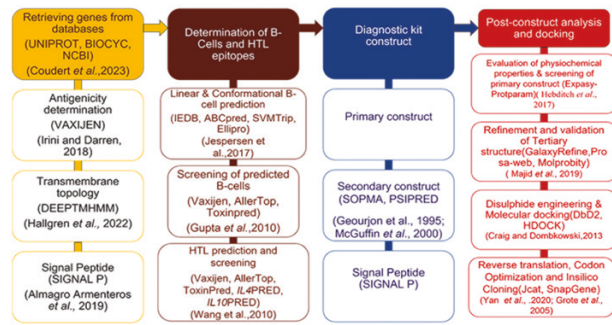


Figure 1. A flowchart of the methodology used in the design of the detection kit.

regions was removed to eliminate false negative detection of the genes in the final detection kit construct.

B cell epitopes in continuous and discontinuous domains

IEDB, ABCpred, and SVMTriP were combined to determine the continuous B cell epitopes. The ABCpred server predicts 15–20 amino acid fixed-length epitopes with a high accuracy of 65%–93%. It uses an artificial neural network algorithm, SVMTriP server uses a support vector machine (SVM) by combining the Tri-peptide similarity and Propensity scores, while IEDB uses a random forest algorithm (Saha and Raghava, 2006; Yao et al., 2012). The epitopes were scrutinized with the following three servers: VaxiJen (Doytchinova and Flower, 2007), AllerTOP (Dimitrov et al., 2014), and ToxinPred (Gupta et al., 2015) for antigenicity, allergenicity, and toxicity. Multiple sequence alignment was done using the Unipro UGENE MAFFT Algorithm (Rose et al., 2019) to determine the consensus epitopes across each gene (Katoh and Toh, 2010; Oladipo et al., 2023b). The ElliPro tool determines the discontinuous B cell epitopes based on their solvent accessibility and electrostatic potential. The application highlights applicant epitope residues and illustrates the predicted antigenic region (Ponomarenko et al., 2008).

IL-4 (interleukin 4) and IL-10 (interleukin 10) inducing epitopes

Predictions of the helper T lymphocytes (HTLs) were made by checking for the major histocompatibility complex II (MHCII) using the web server tool IEDB (Wang et al., 2010). Furthermore, the prediction of *IL-4* and *IL-10* inducing epitopes from HTLs was made using the *IL4*pred and *IL10*pred (Dhanda et al., 2013; Nagpal et al., 2017) servers. Highly antigenic, non-allergenic and non-toxic *IL-4* and *IL-10* inducing epitopes were selected for assemblage.

Multi-epitope construct assemblage and physicochemical functions

The multi-epitope protein sequence was constructed via the top-scoring immunodominant epitopes. B cells, *IL-4* and *IL-10* inducing epitopes with highly antigenic, non-allergenic, and non-toxic properties were selected from each sought-out antigen. To guarantee epitope stability and recognition, these epitopes were bound with a flexible GSGSG linker (Oladipo et

al., 2023a). The Protein-Sol server (Hebditch *et al.*, 2017) evaluated the solubility of the multi-epitope protein using information on *Escherichia coli* protein solubility in the context of a cell-free manufacturing approach. ExPASy-ProtParam (Hennebert *et al.*, 2015) analysed the physical and chemical traits of the protein, particularly its instability index, half-life, grand average hydropathicity (GRAVY), aliphatic index, isoelectric point (pI), atomic composition and molecular weight (MW).

Assemblage secondary and tertiary structure

The alpha and beta helix structure was predicted using the self-optimized prediction method with alignment (SOPMA) server (Deléage, 2017). This tool (Geourjon and Deléage, 1995; McGuffin *et al.*, 2000) uses amino acid sequence to assess the beta sheets, alpha helices, and coils of proteins. In the subsequent phase, the multi-epitope construct underwent 3D homology simulations on the AlphaFold2 server (Mirdita *et al.*, 2022). The AlphaFold 2.0 network precisely forecasts the three-dimensional coordinates of all large atoms for a particular protein using the aligned sequences of homologues and core amino acid sequences as inputs.

Validation and refinement of tertiary models

The best-fit protein model proposed by the AlphaFold2 server was boosted for precision and dependability with the Galaxy Refine server (Lee *et al.*, 2016). The tool recommends five altered models leveraging different optimization parameters (RMSD, MolProbity, GDT-HA, Poor rotamers, Rama, and Clash Score) to improve the projected protein structure. The entire quality score was reviewed using a Z-score by the ProSA-web server (Wiederstein and Sippl, 2007), checking the extent to which they aligned in current protein structures. Furthermore, a Ramachandran map was constructed using the MolProbity service (Williams *et al.*, 2018) to demonstrate the energetically appropriate and disallowed dihedral angles that make up the amino acids psi (ψ) and phi (ϕ).

Protein disulphide engineering

Disulfide by Design (DbD) 2.0 internet server was used for disulphide engineering (Craig and Dombkowski, 2013). The web server identifies residue pairings that can create a disulphide bond if each residue changes to cysteine. This allows the study of protein dynamics, interactions, identification, and mutation of potential residue in the highly mobile region of the sequence to generate new disulphide bonds.

Molecular docking and comparison with genes from other sources

The receptor (Disulfide Engineered structure of the detection kit) and ligand (3D design of each antigenic gene)

molecules were docked using the HDOCK server (Remmert *et al.*, 2012). The server uses a hybrid template-based and template-free docking approach to automatically anticipate their interaction (Yan *et al.*, 2020). Viewing the binding interaction with PDBsum Generate (Laskowski and Thornton, 2022) (Figure 11) showed the protein chains, DNA, ligands, and metal ions and schematic diagrams of their interactions. Sequences of genes collected from additional sources, such as the environment and alfalfa sprouts, were docked to our construct, and the docking score was compared to further ascertain the specificity of the detection kit *in silico*.

Reverse translation, codon optimization, and *in silico* cloning

The Java Codon Adaptation Tool server (JCAT) presented a simple method for adapting codon usage to *E. coli* K12, the cloning vector of interest (Grote *et al.*, 2005). JCAT provides an optimized sequence with codon adaptation index (CAI) and percentage of guanine–cytosine content (GC%). SnapGene (Sarker *et al.*, 2019) was used for *in silico* cloning. Plasmid vectors, peT28a(+) and appropriate restriction sites were used to embed the optimised codon at 5 and 3 –OH positions.

Results

Retrieved sequences and antigenicity score

flbE, *secY*, and *bcsC* (Hartmann *et al.*, 2010; Li *et al.*, 2020) were the genes whose amino acid sequences were obtained in which all of them passed the antigenic threshold of >0.4 for the bacteria model on VaxiJen (Table 1).

Signal peptide and transmembrane topology

The result showed the region of the membrane topology for the gene *bcsC*, *flbE*, and *secY*. The membrane topology result *bcsC* revealed that *bcsC* had an outer region between 797 and 1135 amino acids (aas) with signal peptide, the region of *flbE* outside was between 19 and 133 aas with signal peptide (Figure 2), and *secY* outer region was between 40 and 398 aas with no presence of signal peptide. The signal peptides of *bcsC* and *flbE* were cleaved.

Continuous and discontinuous B cell epitopes

IEDB, SVMtrip, and ABCpred predicted 91 antigenic, non-allergenic, and non-toxic linear B cell epitopes (12 *flbE*, 47 *bcsC*, and 32 *secY*). The consensus genes were also obtained with the Unipro UGENE MAFFT algorithm (Figure 3A). ElliPro predicted seven discontinuous B cell epitopes from the engineered construct (Figure 3B): 16 residues (0.803), 5 residues (0.752), 19 residues (0.744), 18 residues (0.728), 11 residues (0.659), 6 residues (0.656), and 5 residues (0.5).

Table 1. UniProt accession number, gene, function, and antigenic score of the selected genes

Accession number	Gene	Function	Antigenic score
A7MED5	<i>flbE</i>	For flagella production, which is crucial in motility, inflammatory response and ultimately the virulence of the bacteria (Li <i>et al.</i> , 2020)	0.6822
A7MPF4	<i>secY</i>	The gene is used for biofilm formation (Hartmann <i>et al.</i> , 2010)	0.5409
A7MKP2	<i>bcsC</i>	Required for maximal bacterial cellulose synthesis	0.6440

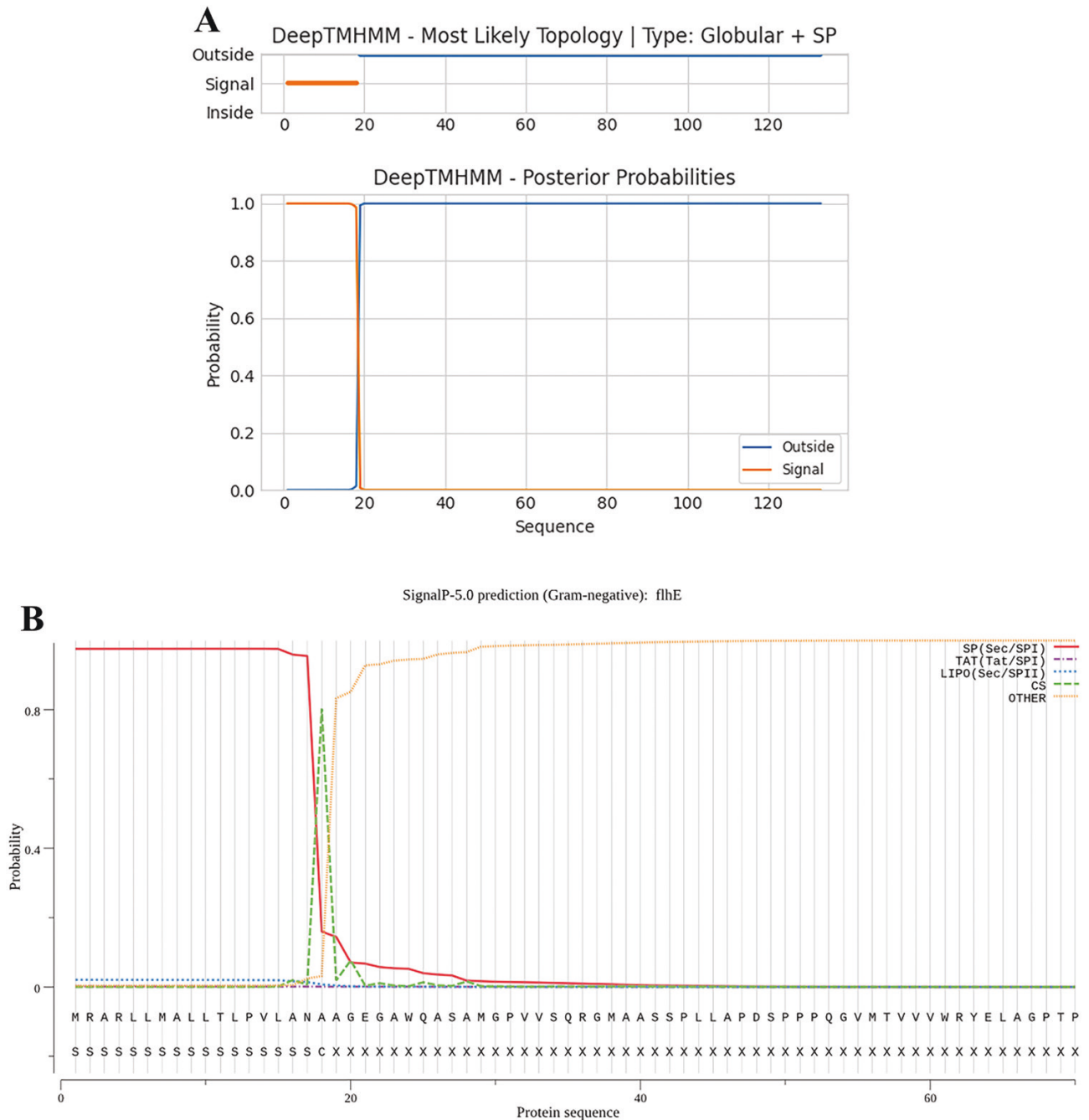


Figure 2. (A) Transmembrane topology (DeepTMHMM). (B) Signal peptide of *flhE* predicted (SignalP).

IL-4 and *IL-10* inducing epitopes

Fifty epitopes were selected from the over 2000 HTL epitopes predicted by the IEDB MHCII server, having passed the antigenicity score (>0.4). Upon further analysis with *IL4*pred and *IL10*pred servers, only *IL-4* and *IL-10* inducing epitopes were carried over (Table 2).

Multi-epitope construct assemblage and physicochemical functions

After thorough prediction and selection to create the multi-epitope sequence, only nine epitopes (linear conformational B cells (LCB) and HTL (*IL-4* and *IL-10* inducing)) emerged.

They were linked with GSGSG, a flexible linker that lengthened the final construct to 162 amino acids (Figure 4).

According to the VaxiJen v2.0 server results, the chimeric model had a high antigenicity index (1.5541). The Protein-Sol server predicted with a score of 0.571 that the design would be soluble when expressed in *E. coli*. The 162 amino acid protein has a molecular weight of 16 584.63 Da and a theoretical pI of 9.34, according to the ExPASy ProtParam web server. The sequence also had eight positively charged residues (Arg+Lys) and six negatively charged residues (Asp+Glu). The generated sequence has an aliphatic index of 71.17, an instability index of 23.26, and a GRAVY score of -0.125 . The protein half-life was discovered to be around 100 h, >20 h,

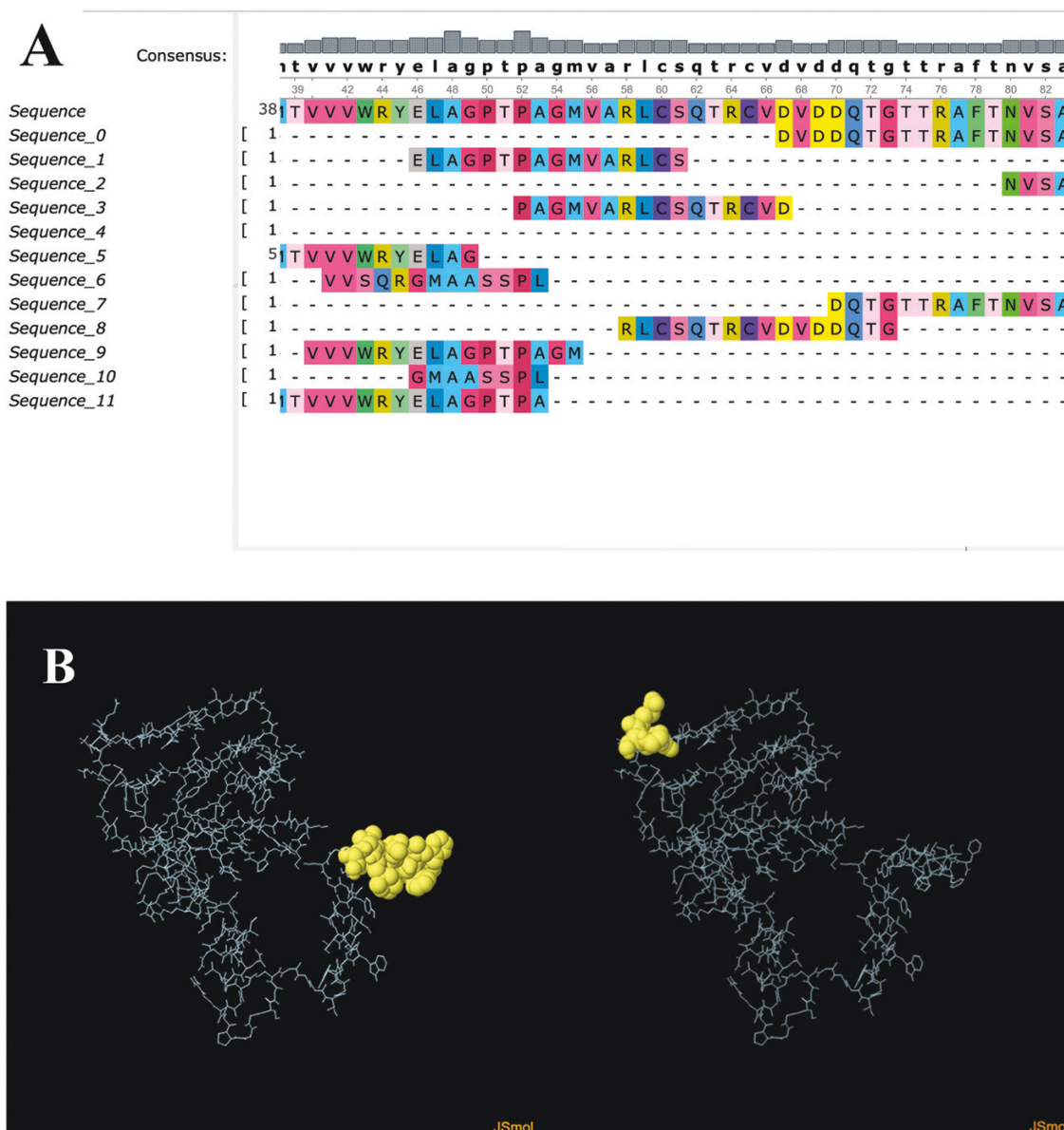


Figure 3. (A) Consensus LCB epitopes for *flhE* from Unipro UGENE. (B) Conformational B cells (2 of 7) predicted by the Ellipro tool of the IEDB online server.

Table 2. *IL-4* and *IL-10* HTL inducing epitopes of *bscC*, *secY*, and *flhE*

Gene	Allele	Epitopes	Antigenic score (VaxiJen)	IL4 inducer (<i>IL4</i> pred)	IL10 inducer (<i>IL10</i> pred)	Allergenic (ALLERTOP)	Toxic (TOXINPRED)
<i>bscC</i>	HLA-DRB4*01:01	RDEAIRQMQLDARA	0.9268	Yes	Yes	No	No
	HLA-DRB4*01:01	EAIRQMQLDARAPG	0.6747	Yes	Yes	No	No
<i>secY</i>	HLA-DRB1*07:01	GGTGWNWLTITISLYL	1.0979	Yes	Yes	No	No
	HLA-DRB1*15:01	LYVLLYASAIIFCF	1.194	No	Yes	No	No
<i>flhE</i>	HLA-DRB3*01:01	VVVWRYELAGPTPAG	0.9552	Yes	No	No	No

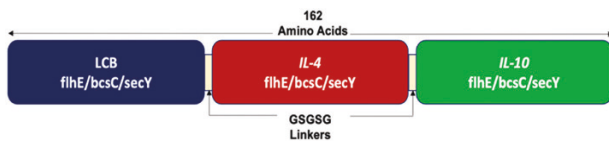


Figure 4. Primary construct of the peptide based kit.

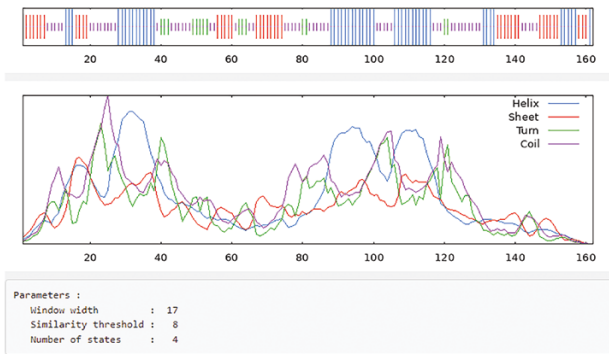


Figure 5. Secondary structure, SOPMA and PSIPRED.

and >10 h in human reticulocytes (*in vitro*), yeast (*in vivo*), and *E. coli* (*in vivo*), respectively.

Secondary and tertiary structure of the assemblage

The SOPMA server provided the following details: 58 (35.80%) random coil, 41 (25.31%) extended strands, 15 (9.26%) beta-turn, and 48 (29.63%) alpha helix. The secondary structure's most predominant areas, as the PSIPRED server shows, were coil regions, followed by helix and strand structures. [Figure 5](#) presents the illustration of the PSIPRED server's results. Based on their rankings, the AlphaFold2 server displayed five tertiary structure models of the multi-epitope protein. Therefore, using UCSF ChimeraX, the model with rank one was modelled as the most appropriate model ([Pettersen et al., 2021](#)).

Tertiary model refinement and validation

The GalaxyRefine server for the designed construct predicted five refined models. Model 1 was selected as the most appropriate model for further study because it satisfied the qualifying criteria of GDTHA of 0.8858, RMSD of 0.603, MolProbity of 1.433, Clash score of 3.4, Poor rotamers of 0.0, and Rama favoured of 95.6. The improved model ([Figure 6](#)), which is the most appropriate, has a Z-score of -3.41 based on the ProSA-web outcome ([Figure 7](#)). Approximately 95.6% of all residues were found in the preferred areas, whereas 97.5% were found in the allowed regions according to the MolProbity Ramachandran plot analysis of the improved model ([Figure 8B](#)). As shown in [Figure 8A](#), the crude Ramachandran analysis revealed that 63.8% of the residues were in the favoured regions and 78.8% were in the allowed regions.

Protein disulphide engineering

Only one pair of amino acids, ALA 120–GLY 123, was found to satisfy the criteria for the formation of disulphide bonds after residues were screened with DbD2 using the requirements of a chi3 value of -87° or $+97^\circ$ and an energy value of

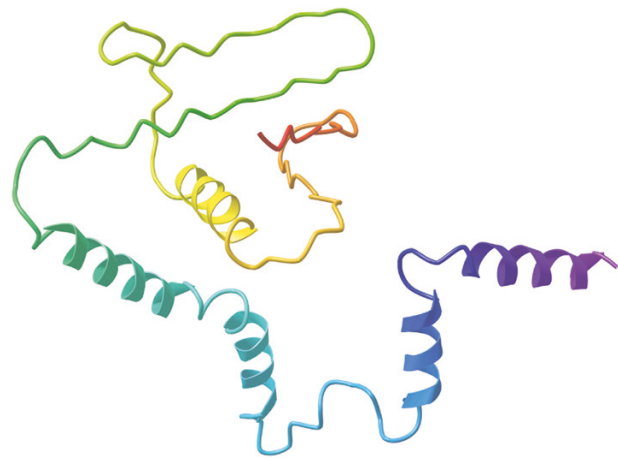


Figure 6. Refined tertiary protein structure, GalaxyRefine.

2.5 B-factor. These residues were then switched out for cysteine residues ([Figure 9](#)).

Molecular docking and comparison with genes from other sources

The HDOCK server provided the docked model ranked 1–10; each had a docking score <-200 and a confidence score ranging from 0.9000 to 1.000 ([Table 3](#)) ([Figure 10](#)).

In silico cloning

The improved sequences exhibited a CAI value of 1.0 and the percentage of cytosine and guanine nucleotides (CG%) of 55.35%. In contrast, that of *E. coli* (strain K12) was 50.73%. For the *in silico* cloning, *TatI* (5' start) and *BstAPL* (3' end) were the restriction sites used. Upon inserting the fragment into the pET28a(+) vector, a cloned structure of 2683 bp was obtained ([Figure 12](#)).

Discussion

Cronobacter sakazakii, an emerging neonatal pathogen, has attracted considerable attention globally. It has been associated with outbreaks of life-threatening septicaemia, necrotizing enterocolitis, and meningitis in neonates and infants. The xerotolerant pathogen uses a variety of genes to ensure its survival in dry conditions; therefore, it can be detected in stored PIF even after 2.5 years of storage ([Srikumar et al., 2019](#); [Elkhawaga et al., 2020](#)). Currently, the detection methods used for *C. sakazakii* either have a longer turnaround time, require specialized equipment handled by highly trained specialists, or are expensive. Further research is needed to develop new techniques to create a simple, compact, portable device to facilitate early and rapid outbreak surveillance. Therefore, designing a highly sensitive and specific detection kit via immunoinformatics will be beneficial.

Studies have shown that the ability of *C. sakazakii* to survive under desiccated stress can be attributed to its ability to form biofilms. It helps the organism firmly adhere to the surface of equipment and packaging materials, increasing the possibility of contamination of PIF with this pathogen ([Aly et al., 2019](#)). *bcsC* (cellulose biosynthesis), *flhE* (flagellar structure or biosynthesis), and *secY* genes have been found to play significant roles in biofilm formation ([Hartmann et al., 2010](#);

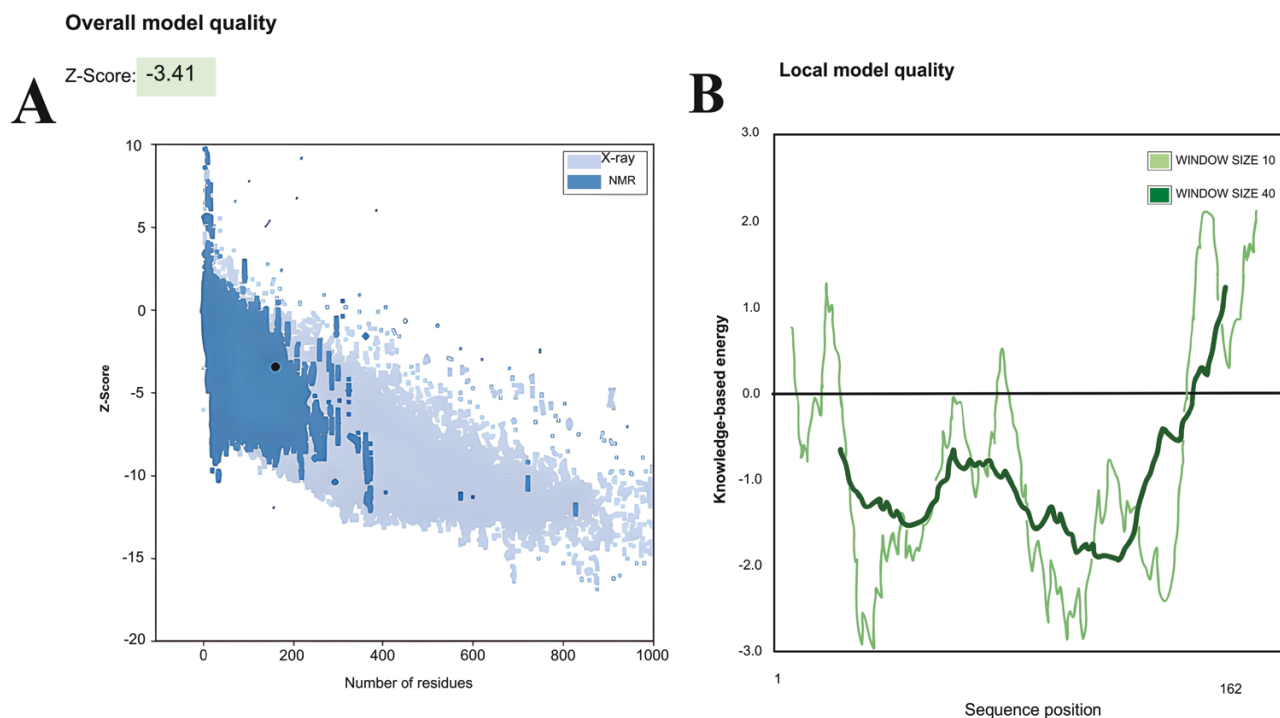


Figure 7. Validation of the refinement process with a Z-score of -3.41 based on the ProSA-web.

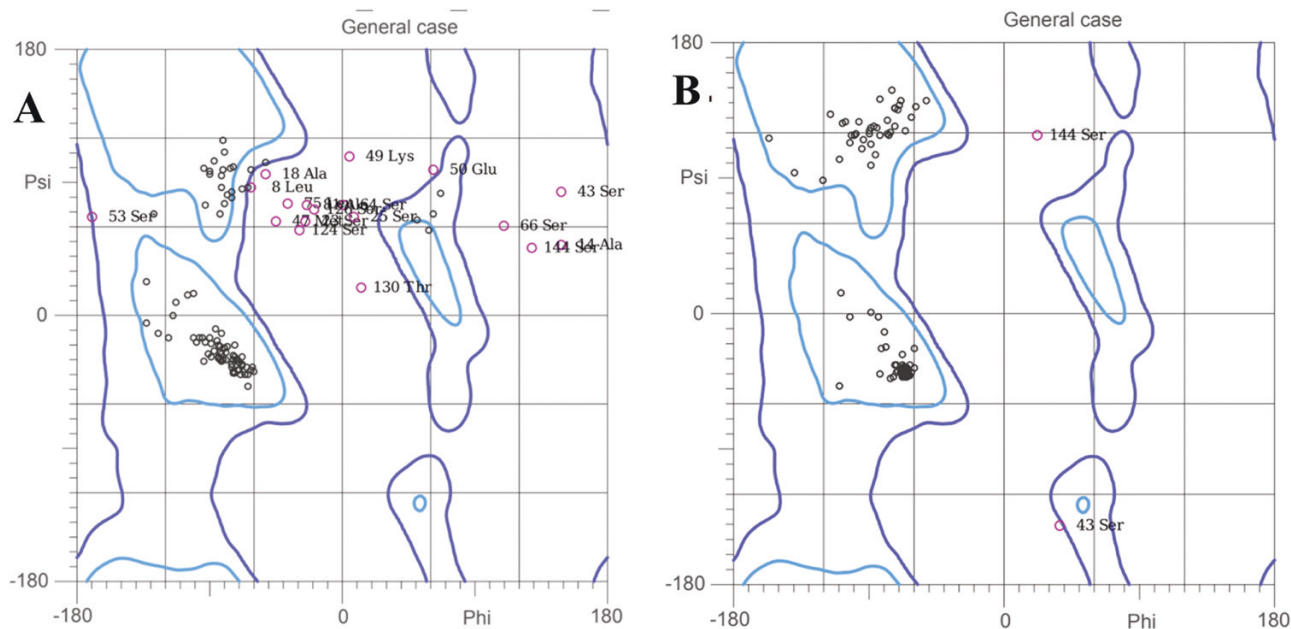


Figure 8. MolProbity server's Ramachandran plot analysis of (A) the crude model and (B) the refined model.

Li *et al.*, 2020). They were screened to choose potential candidates for the detection kit construct. Using VaxiJen, the protein sequences derived from the genes were subjected to antigenicity screening at a threshold of 0.4, where proteins with exact or higher antigenicity imply that they can elicit an immune response and bind to specific antibodies (Doytchinova and Flower, 2007). Subsequently, these antigenic genes were subjected to a DeepTMHMM online server to analyse their transmembrane topology. A more significant portion of the protein sequences of these three genes were found outside the

membrane. As Yao *et al.* (2022) described, using antigenic proteins with epitopes in their outer membrane sequence will increase exposure and easy detection for antibody-antigen interaction without the need for a lysis buffer.

Continuous and discontinuous B cell epitopes were predicted with three servers (IEDB, SVMTrip, and ABCpred). Unipro UGENE was used to obtain a consensus for the epitopes of each gene. For the detection kit design, the selected peptides had 12–20 amino acid residues to get a reasonable molecular weight. Further antigenicity, allergenicity and

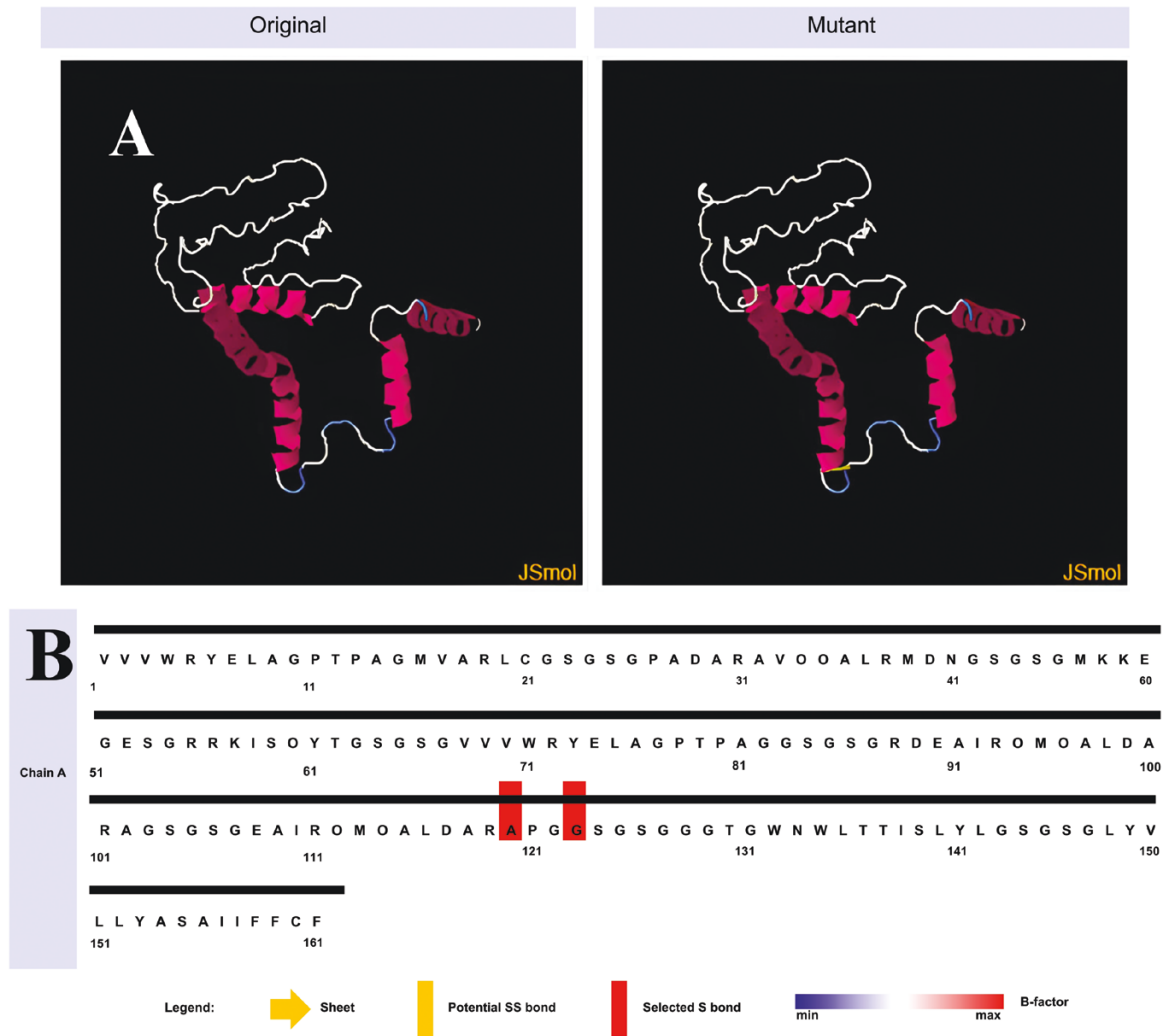


Figure 9. Disulfide Design results in (A) the original and (B) the mutant model.

Table 3. Docking results of multi-epitope peptide and genes of *Cronobacter sakazakii* from PIF and other sources

Source	Gene	Docking score	Conference score
PIF	<i>bcsC</i>	-328.48	0.9726
	<i>secY</i>	-369.53	0.9878
	<i>flbE</i>	-260.63	0.9014
Environment	<i>ompA</i>	-282.06	0.9335
Alfalfa sprouts	<i>flbE</i>	-263.95	0.9071

toxicity screening were conducted on the epitopes. Epitopes with the highest antigenicity score that could induce *IL-4* and *IL-10* were selected. *IL-4* helps B cells differentiate and switch between different types of antibodies, whereas *IL-10* is crucial for B cells' survival, growth and generation of antibodies, as supported by Moore et al. (2001). The predicted epitopes that passed all the necessary screenings were linked with the GSGSG linker to increase construct folding and biological activity. As a result, the primary construct comprised 162 amino acid residues with a high antigenic score of 1.5441. The ElliPro tool of the IEDB server was used to predict seven

discontinuous B cell epitopes. These epitopes are essential for determining the quality of interactions between antibodies and antigens. The protein-Sol server yielded a solubility coefficient of 0.5711, indicating that the construct is soluble. Other physicochemical features, such as theoretical pI (9.34), instability index (23.26), and aliphatic index (71.17) confer stability and the ability of the construct to withstand thermal denaturation. At the same time, a GRAVY of -0.125 shows it is hydrophilic. According to Shams et al. (2021), such theoretical pI is advantageous for ion-exchange chromatography and isoelectric focusing techniques. The multi-epitope construct had an estimated half-life in yeast to be >20 h; in *E. coli*, the estimated half-life was >10 h *in vivo* and the half-life was 100 h in mammalian reticulocytes *in vitro*. The GC content and CAI value are essential for successful protein expression in the prokaryotic host. Typically, the GC content falls within the 35%–65% range, while the CAI value ranges from 0.8 to 1.0. In this study, the JCAT online server was employed to enhance the GC content and CAI value of the multi-epitope protein to 55.35% and 1.0, respectively, to improve its expression in the *E. coli* K12 strain. The GC content of the *E. coli* K12 strain was measured to be 50.73%. The protein model underwent reverse translation and codon adaptation. This modified DNA sequence was then successfully ligated

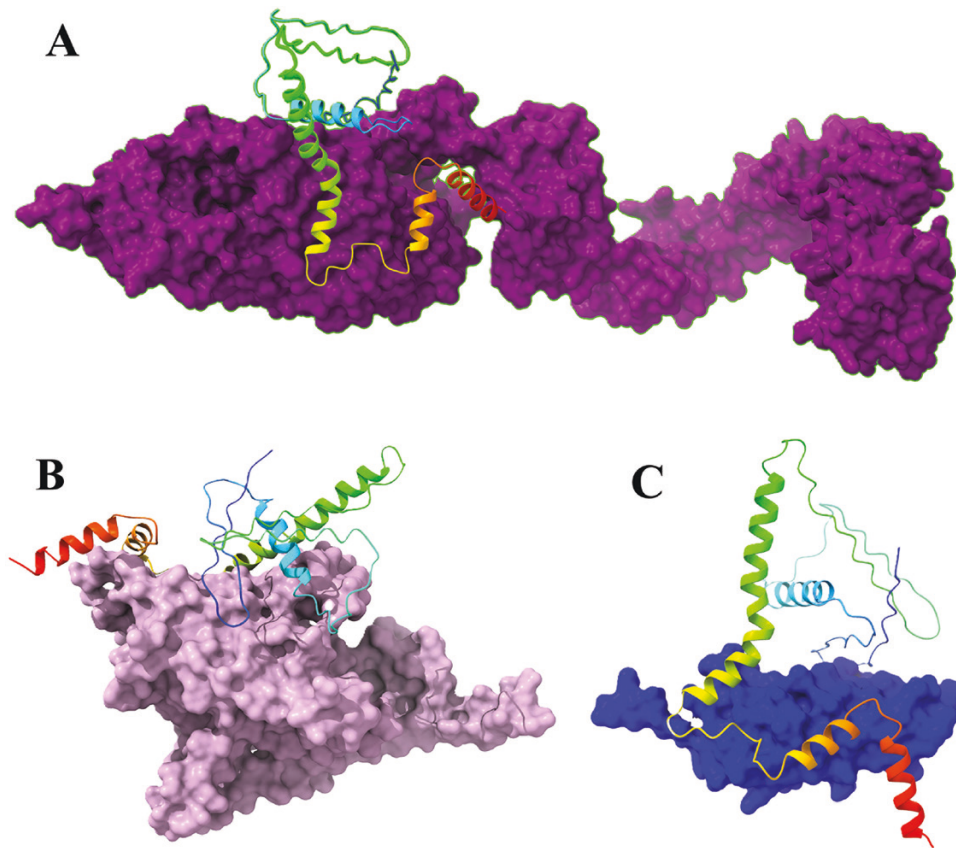


Figure 10. HDock model viewed with UCSF ChimeraX. (A) Construct and *bcsC*. (B) Construct and *secY*. (C) Construct and *flhE*.

with the pET 28a(+) vector using the SnapGene software for cloning purposes. To precisely identify and cleave the DNA sequences in a controlled manner, restriction enzymes BstAPL and *TatI* were used. The *E. coli* expression system is a better option due to its low cost and ease of bacterial cultivation (Kamionka, 2011). The T7 promoter, frequently used to control the gene expression of recombinant proteins, is included in this vector, making it an excellent system for producing large quantities of the required peptide (Safavi *et al.*, 2019). Additionally, it has proper restriction enzyme sites.

SOPMA server analysis of the protein secondary structure shows that the percentage of random coils, extended strands, beta (β) turn, and alpha (α) helices were 58 (35.80%), 41 (25.31%), 15 (9.26%), and 48 (29.63%), respectively. The tertiary construct of the epitope was predicted using the AlphaFold 2.0 online server and refined using GalaxyRefine. Five models were constructed, but the best model was model 1, which had a better Rama-favoured score (95.6) and GDT-HA score (0.8858). The ProSA-web and MolProbity servers were utilised to validate the quality of the refining process. The ProSA-web results demonstrated a significant improvement, indicated by a Z-score of -3.41 . Additionally, Ramachandran analysis by MolProbity showed that the refined model had 95.6% and 97.5% of its amino acids in the favoured region of the Ramachandran plot, which indicates that the refined model is of high quality. This is compared to the Ramachandran analysis of the crude model, which had 63.8% and 78.8% of its residue in the favoured and allowed areas, respectively. Therefore, the crude model has lower quality than the refined model. The result from DbD2 showed that the sequence was mutated at ALA120 and GLY123

residue lengths to increase the thermal stability of the epitope as well as improve its interaction, as explained earlier by Gao *et al.* (2020).

Subsequently, by employing the HDock online server, the construct, which passed through refinement, validation and disulphide engineering processes, was docked against the three antigenic genes whose source was PIF. It was examined that the docking score was <-200 . A similar technique was carried out for identifying antigenic genes from other sources, such as the environment and alfalfa sprouts. However, their binding scores were lower than those of the PIF isolates, except for that of *flhE*, which was close. This suggests that the primary construct has favourable interaction or stronger binding affinity with PIF isolates than those from the environment and alfalfa sprouts, except for *flhE*.

Conclusions

Using bioinformatics and computational methods to design a detection kit for detecting *C. sakazakii* holds excellent potential for creating a new detection tool for this foodborne pathogen. It can be concluded that the final construct and clone have successfully met the *in silico* design requirements for the development of a lateral flow kit, which has the potential to provide quick and effective detection of *C. sakazakii*, based on the findings from epitope peptide analysis Z-score, Ramachandran plot analysis, cloning, and docking. This could enhance treatment options and accelerate decision-making, ultimately decreasing death rates. However, it is critical to recognize that further confirmation using *in vitro* and *in vivo* methods is needed.

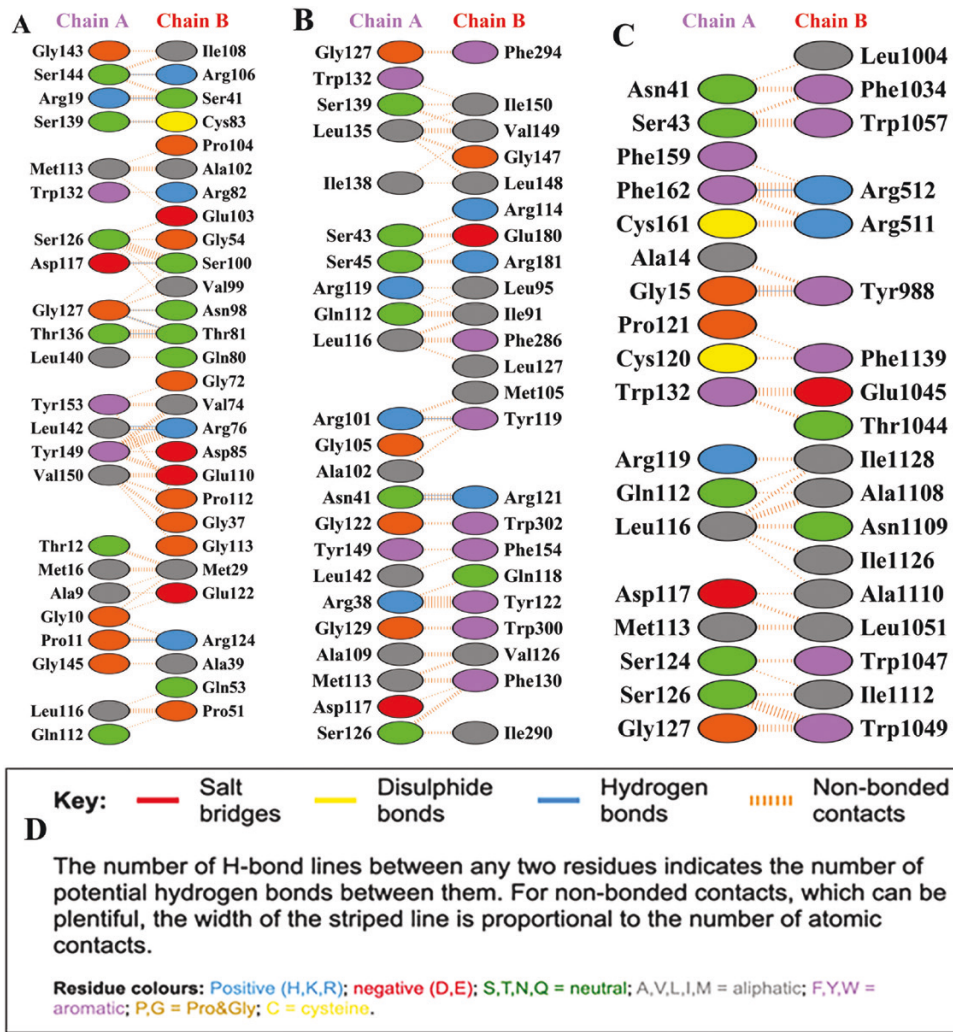


Figure 11. Interactions between construct multi-epitope (chain A) and the genes (*flhE*, *secY*, and *bcsC*) in chain B as shown in PDBsum.

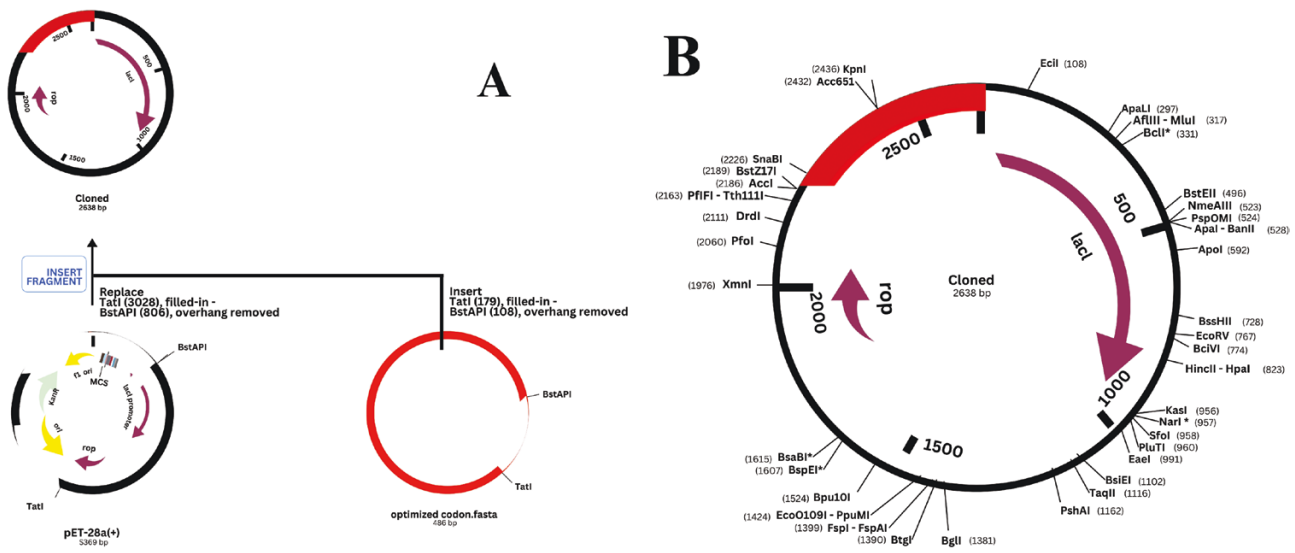


Figure 12. *In silico* restriction cloning of the sequence of the multi-epitope construct into pET28a(+) expression vector.

Acknowledgement

We expressed our gratitude to the whole administration and team of the Helix Biogen Institute in Ogbomoso, Nigeria, for their cooperation with the technical aspects of the project.

Funding

The research received no external funding.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Elijah K. Oladipo and Helen Onyeaka: Conceptualization, supervision, editing; Elijah K. Oladipo and Temitope A. Oyedepo: Original draft preparation, methodology; Olukayode I. Obisanya, Victoria O. Owoeye, and Oyinlola G. Shittu: Writing and reviewing; Mautin G. Adamitonde, Emmanuel C. Ikwuka, and Solomon O. Ojewale: Methodology, data Analysis; Amudatu A. Adedokun, Adewale J. Tijani, and Feranmi A. Adedokun: Software, investigation.

References

- Aly, M. A., Reimhult, E., Kneifel, W., *et al.* (2019). Characterization of biofilm formation by *Cronobacter* spp. isolates of different food origin under model conditions. *Journal of Food Protection*, 82(1): 65–77.
- Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., *et al.* (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature biotechnology*, 37(4): 420–423.
- Cai, X. Q., Yu, H. Q., Ruan, Z. X., *et al.* (2013). Rapid detection and simultaneous genotyping of *Cronobacter* spp. (formerly *Enterobacter sakazakii*) in powdered infant formula using real-time PCR and high resolution melting (HRM) analysis. *PLoS One*, 8(6): e67082.
- Caubilla-Barron, J., Hurrell, E., Townsend, S., *et al.* (2007). Genotypic and phenotypic analysis of *Enterobacter sakazakii* strains from an outbreak resulting in fatalities in a neonatal intensive care unit in France. *Journal of Clinical Microbiology*, 45(12): 3979–3985.
- Craig, D. B., Dombkowski, A. A. (2013). Disulfide by Design 20: a web-based tool for disulfide engineering in proteins. *BMC Bioinformatics*, 14(1): 1–7.
- Deléage, G. (2017). ALIGNSEC: viewing protein secondary structure predictions within large multiple sequence alignments. *Bioinformatics*, 33(24): 3991–3992.
- Dhanda, S. K., Gupta, S., Vir, P., *et al.* (2013). Prediction of IL4 inducing peptides. *Clinical and Developmental Immunology*, 2013: 1–9.
- Dimitrov, I., Bangov, I., Flower, D. R., *et al.* (2014). AllerTOP v. 2—A server for in silico prediction of allergens. *Journal of Molecular Modeling*, 20: 1–6.
- Doytchinova, I. A., Flower, D. R. (2007). VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*, 8(1): 1–7.
- Elkhwaga, A. A., Hetta, H. F., Osman, N. S., *et al.* (2020). Emergence of *Cronobacter sakazakii* in cases of neonatal sepsis in upper Egypt: first report in North Africa. *Frontiers in Microbiology*, 11: 215.
- Fei, P., Man, C., Lou, B., *et al.* (2015). Genotyping and source tracking of *Cronobacter sakazakii* and *C. malonaticus* isolates from powdered infant formula and an infant formula production factory in China. *Applied and Environmental Microbiology*, 81(16): 5430–5439.
- Fei, P., Jing, H., Ma, Y., *et al.* (2022). *Cronobacter* spp. in commercial powdered infant formula collected from nine provinces in China: prevalence, genotype, biofilm formation, and antibiotic susceptibility. *Frontiers in Microbiology*, 13: 900690.
- Gao, X., Dong, X., Li, X., *et al.* (2020). Prediction of disulfide bond engineering sites using a machine learning method. *Scientific Reports*, 10(1): 1–9.
- Geourjon, C., Deleage, G. (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Computer Applications in the Biosciences*, 11(6): 681–684.
- Grote, A., Hiller, K., Scheer, M., *et al.* (2005). JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Research*, 33: W526–W531.
- Gupta, S., Kapoor, P., Chaudhary, K., *et al.* (2015). Peptide toxicity prediction. *Methods in Molecular Biology*, 128: 143–157.
- Hallgren, J., Tsirigos, K. D., Pedersen, M. D., *et al.* (2022). DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. *BioRxiv*, in press.
- Hariri, S., Joseph, S., Forsythe, S. J. (2013). *Cronobacter sakazakii* ST4 strains and neonatal meningitis, United States. *Emerging Infectious Diseases*, 19(1): 175–177.
- Hartmann, I., Carranza, P., Lehner, A., *et al.* (2010). Genes involved in *Cronobacter sakazakii* biofilm formation. *Applied and Environmental Microbiology*, 76(7): 2251–2261.
- Healy, B., Cooney, S., O'Brien, S., *et al.* (2010). *Cronobacter (Enterobacter sakazakii)*: an opportunistic foodborne pathogen. *Foodborne Pathogens and Disease*, 7(4): 339–350.
- Hebditch, M., Carballo-Amador, M. A., Charonis, S., *et al.* (2017). Protein-Sol: a web tool for predicting protein solubility from sequence. *Bioinformatics*, 33(19): 3098–3100.
- Hennebert, E., Maldonado, B., Ladurner, P., *et al.* (2015). Experimental strategies for the identification and characterization of adhesive proteins in animals: a review. *Interface Focus*, 5(1): 20140064.
- Iversen, C., Lehner, A., Mullane, N., *et al.* (2007). The taxonomy of *Enterobacter sakazakii*: proposal of a new genus *Cronobacter* gen. nov. and descriptions of *Cronobacter sakazakii* comb. nov., *Cronobacter sakazakii* subsp. *sakazakii*, comb. nov., *Cronobacter sakazakii* subsp. *malonaticus* subsp. nov., *Cronobacter turicensis* sp. nov., *Cronobacter muytjensii* sp. nov., *Cronobacter dublinensis* sp. nov. and *Cronobacter* genomospecies 1. *BMC Evolutionary Biology*, 7: 1–11.
- Kamionka, M. (2011). Engineering of therapeutic proteins production in *Escherichia coli*. *Current Pharmaceutical Biotechnology*, 12(2): 268–274.
- Katoh, K., Toh, H. (2010). Parallelisation of the MAFFT multiple sequence alignment program. *Bioinformatics*, 26(15): 1899–1900.
- Laskowski, R. A., Thornton, J. M. (2022). PDBsum extras: SARS-Cov-2 and AlphaFold models. *Protein Science*, 31(1): 283–289.
- Lee, G. R., Heo, L., Seok, C. (2016). Effective protein model structure refinement by loop modeling and overall relaxation. *Proteins: Structure, Function, and Bioinformatics*, 84: 293–301.
- Li, C., Sun, X., Niu, B., *et al.* (2020). Exopolysaccharide-related gene *bcsG* affects biofilm formation of *Cronobacter* spp. *International Dairy Journal*, 111: 104844.
- McGuffin, L. J., Bryson, K., Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics*, 16(4): 404–405.
- McMullan, R., Menon, V., Beukers, A. G., *et al.* (2018). *Cronobacter sakazakii* infection from expressed breast milk, Australia. *Emerging Infectious Diseases*, 24(2): 393–394.
- Mirdita, M., Schütze, K., Moriwaki, Y., *et al.* (2022). ColabFold: making protein folding accessible to all. *Nature Methods*, 19(6): 679–682.
- Moore, K. W., de Waal Malefyt, R., Coffman, R. L., *et al.* (2001). Interleukin-10 and the interleukin-10 receptor. *Annual review of immunology*, 19(1): 683–765.
- Nagpal, G., Usmani, S. S., Dhanda, S. K., *et al.* (2017). Computer-aided designing of immunosuppressive peptides based on IL-10 inducing potential. *Scientific Reports*, 7(1): 42851.
- Oladipo, E. K., Jimah, E. M., Irewolede, B. A., *et al.* (2023a). Immuno-informatics design of multi-epitope peptide for the diagnosis of *Schistosoma haematobium* infection. *Journal of Biomolecular Structure and Dynamics*, 41(14): 6676–6683.

- Oladipo, E. K., Oluwasegun, J. A., Oladunni, T. D., et al. (2023b). Immunoinformatics aided design of a peptide-based kit for detecting *Escherichia coli* O157:H7 from food sources. *Journal of Food Safety*, 43(5): e13073.
- Pettersen, E. F., Goddard, T. D., Huang, C. C., et al. (2021). UCSF ChimeraX: structure visualisation for researchers, educators, and developers. *Protein Science*, 30(1): 70–82.
- Ponomarenko, J., Bui, H. H., Li, W., et al. (2008). ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics*, 9(1): 1–8.
- Remmert, M., Biegert, A., Hauser, A., et al. (2012). HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nature Methods*, 9(2): 173–175.
- Rose, R., Golosova, O., Sukhomlinov, D., et al. (2019). Flexible design of multiple metagenomics classification pipelines with UGENE. *Bioinformatics*, 35(11): 1963–1965.
- Safavi, A., Kefayat, A., Sotoodehnejadnematlahi, F., et al. (2019). Production, purification, and *in vivo* evaluation of a novel multiepitope peptide vaccine consisted of immunodominant epitopes of SYCP1 and ACRBP antigens as a prophylactic melanoma vaccine. *International Immunopharmacology*, 76: 105872.
- Saha, S., Raghava, G. P. S. (2006). Prediction of continuous B-cell epitopes in an antigen using recurrent neural networks. *Proteins: Structure, Function, and Bioinformatics*, 65(1): 40–48.
- Sarker, A., Rathore, A. S., Gupta, R. D. (2019). Evaluation of scFv protein recovery from *E. coli* by *in vitro* refolding and mild solubilization process. *Microbial Cell Factories*, 18(1): 5.
- Schoch, C. L., Ciuffo, S., Domrachev, M., et al. (2020). NCBI taxonomy: a comprehensive update on curation, resources and tools. *Data-base*, 2020: baaa062.
- Shams, M., Nourmohammadi, H., Asghari, A., et al. (2021). Construction of a multi-epitope protein for human *Toxocara canis* detection: Immunoinformatics approach multi-epitope construct for *T. canis* serodiagnosis. *Informatics in Medicine Unlocked*, 26: 100732.
- Srikumar, S., Cao, Y., Yan, Q., et al. (2019). RNA sequencing-based transcriptional overview of xerotolerance in *Cronobacter sakazakii* SP291. *Applied and Environmental Microbiology*, 85(3): e01993–e01918.
- Wang, P., Sidney, J., Kim, Y., et al. (2010). Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC bioinformatics*, 11: 1–12.
- Wiederstein, M., Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, 35: W407–W410.
- Williams, C. J., Headd, J. J., Moriarty, N. W., et al. (2018). MolProbity: more and better reference data for improved all-atom structure validation. *Protein Science*, 27(1): 293–315.
- Yao B., Zhang, L., Liang, S., et al. (2012). SVMTriP: a method to predict antigenic epitopes using a support vector machine to integrate tri-peptide similarity and propensity. *Nature Protocols*, 15(5): 1829–1852.
- Yan, Y., Tao, H., He, J., et al. (2020). The HDock server for integrated protein–protein docking. *Nature Protocols*, 15(5): 1829–1852.
- Yao, M., Guo, X., Wu, X., et al. (2022). Evaluation of the combined use of major outer membrane proteins in the serodiagnosis of *Brucella*. *Infection and Drug Resistance*, 15: 4093–4100.