

Immunoinformatics Assisted Design of a Multi-Epitope Kit for Detecting *Cronobacter sakazakii* in Powdered Infant Formula (PIF)

Elijah K. Oladipo^{1,3,8}, Olukayode I. Obisanya^{2,3}, Victoria O. Owoeye^{3,4}, Oyinlola G. Shittu⁵, Mautin G. Adamitonde⁶, Emmanuel C. Ikwuka⁷, Solomon O Ojewale⁷, Adewale J. Tijani⁹ and Feranmi A. Adedokun⁹, Amudatu A. Adedokun¹⁰, Temitope A. Oyedepo¹¹ Helen Onyeaka^{12*}

1. Department of Microbiology, Laboratory of Molecular Biology, Immunology and Bioinformatics, Adeleke University, Ede, Osun State, Nigeria.
2. Department of Medical Laboratory Science, College of Medicine, University of Lagos.
3. Genomics Unit, Helix Biogen Institute, Ogbomoso, Oyo State, Nigeria.
4. Department of Microbiology, Obafemi Awolowo University, Ile-Ife, Osun state, Nigeria.
5. Department of Physiology, Ladoke Akintola University of Technology, Ogbomoso, Oyo State, Nigeria.
6. Department of Human Anatomy, Ladoke Akintola University of Technology, Ogbomoso, Oyo State, Nigeria.
7. Department of Biotechnology, Federal University of Technology, Akure, Ondo State, Nigeria
8. Department of Microbiology, Ladoke Akintola University of Technology, Ogbomoso, Oyo State, Nigeria.
9. Department of Biochemistry, Ladoke Akintola University of Technology, Ogbomoso, Oyo State, Nigeria.
10. Medical Laboratory Science Department, Fountain University, Osogbo, P. M. B. 4491, Osun State, Nigeria
11. Department of Biochemistry, Adeleke University, Ede, Osun State, Nigeria
12. School of Chemical Engineering. University of Birmingham, Edgbaston, Birmingham B12 2TT, UK

*Corresponding author: Helen Onyeaka (h.onyeaka@bham.ac.uk)

© 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

Abstract

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, septicemia and necrotising enterocolitis among infants worldwide in association with powdered infant formula (PIF). In the present study, a peptide-based kit was designed with a bioinformatic technique to rapidly identify *Cronobacter sakazakii* in powdered infant formula (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 disulfide 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. 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 kilo Dalton (kDa), and a Z-score of -3.41. The multi-epitope peptide construct could be a conceptual framework for creating a *Cronobacter 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* (*C. 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: *Cronobacter sakazakii*, *Cronobacter malonaticus*, *Cronobacter turicensis*, *Cronobacter muytjensii*, *Cronobacter condimenti*, *Cronobacter universalis*, and *Cronobacter dublinensis* (Iversen *et al.*, 2007). Several strains of *Cronobacter sakazakii*, *C. malonaticus*, *C. turicensis* and *C. dublinensis* are desiccation-resistant and persistent in dried products such as powdered infant formula (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 *Cronobacter sakazakii* infections have been reported to affect newborns, children, and people with weakened immune systems (Healy *et al.*, 2010). Numerous countries reported nosocomial outbreaks in NICUs (neonatal intensive care units) (Caubilla-Barron *et al.*, 2007; Hariri *et al.*, 2013) with high mortality rates ranging from 40 – 80% (Friedemann *et al.*, 2007).

The current methods used to detect and identify *C. sakazakii* include culture and biochemical-

based methods, PCR or gene probe assays. These methods require highly skilled personnel and expensive equipment or longer turnaround time (TAT); 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 has also been a need for more consistency between results obtained with different detection and identification methods (Hassan *et al.*, 2007); the current United States Food and Drug Administration (FDA) 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 insilico approach to design a detection kit for specific targets for pathogen detection (Schoch *et al.*, 2020; Oladipo *et al.*, 2023). However, no such study has been done for *Cronobacter species*. In this study, bioinformatics tools were extensively utilised for the insilico design of a detection kit to effectively detect *Cronobacter sakazakii* in Powdered Infant Formula (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.*, 2022). Using the Vaxijen 2.0 Bacteria model, it was predicted that these genes would trigger a response in the immune system (Doytchinova & Flower, 2007).

Vaxijen generates protein antigenicity without using alignment and emphasises the physical and chemical features of the identified alternative (Irimi & Darren, 2018).

Signal Peptide and Transmembrane Topology

DeepTMHMM Server (Hallgren *et al.*, 2022) predicted 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 acids sequence that falls on these regions was removed to eliminate false negative detection of the genes in the final detection kit construct.

B-Cells 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 acids fixed-length epitopes with a high accuracy of 65-93%. It uses an artificial neural network (ANN) 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 & Raghava, 2006; Yao *et al.*, 2012). The epitopes were scrutinised with these three servers: VaxiJen (Irimi & Darren, 2018), AllerTOP (Dimitrov *et al.*, 2014), and ToxinPred (Gupta *et al.*, 2010) for antigenicity, allergenicity, and toxicity. Multiple sequence alignment was done using the Unipro UGENE MAFFT Algorithm (Rose *et al.*, 2009) to determine the consensus epitopes across each gene (Katoh & Stanley, 2013; Oladipo *et al.*, 2023). 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 interleukin four and interleukin ten inducing epitopes from HTLs was made using the *IL4*pred and *IL10*pred (Raghava, 2014; 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.*, 2022). The Protein-Sol server (Hebditch *et al.*, 2017) evaluated the solubility of the multi-epitope protein using information for *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 hydrophobicity (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-optimised prediction method with alignment (SOPMA) server (Deléage, 2017). This tool (Geourjon *et al.*, 1995; McGuffin *et al.*, 2000) uses the 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 (Jumper *et al.*, 2021).

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 optimisation 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 & 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 Disulfide Engineering

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

Molecular Docking & 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.*, 2011). 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 & Thornton, 2022) 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 to adapt the codon usage to *E. coli* K12, our cloning vector of interest (Grote *et al.*, 2005). JCAT provides an optimised sequence with codon adaptation index (CAI) and percentage of GC content (GC%) (Morteza *et al.*, 2021). 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 in 5' and 3' –OH positions.

RESULTS

Retrieved Sequences and Antigenicity Score

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

Table 1. UniProt Accession number, Gene, Function and Antigenic Score of selected genes

Accession Number	Gene	Function	Antigenic Score
A7MED5	<i>flhE</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

Signal Peptide and Transmembrane Topology

The result showed the region of the membrane topology for the gene listed: *bcsC*, *flhE*, *secY*. For the membrane topology result of *bcsC* had an outer region between 797-1135 amino acids (aas) with signal peptide; the region of *flhE* outside was between 19-133 aas with signal peptide (Fig 2), and *SecY* outer region was between 40-398 aas with no presence of signal peptide. The signal peptides of *bcsC* and *flhE* were cleaved.

Continuous and Discontinuous B-cell epitopes

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

IL-4 and IL-10 Inducing Epitopes

Fifty epitopes were selected from the over 2,000 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).

Table 2. *IL-4 & IL-10* HTL Inducing Epitopes of *BscC*, *SecY* and *FlhE*

Gene	Allele	Epitopes	Antigenic Score (VAXIJEN)	IL4 Inducer (IL4PRED)	IL10 Inducer (IL10PRED)	Allergenic (ALLERTOPO)	Toxic (TOXINPREDED)
BscC	HLA-DRB4*01:01	RDEAIRQMQUALD ARA	0.9268	YES	YES	NO	NO
	HLA-DRB4*01:01	EAIRQMQUALDAR APG	0.6747	YES	YES	NO	NO
SecY	HLA-DRB1*07:01	GGTGWNWLTTIS LYL	1.0979	YES	YES	NO	NO
	HLA-DRB1*15:01	LYVLLYASAIIFFC F	1.194	NO	YES	NO	NO
FlhE	HLA-DRB3*01:01	VVVWRYELAGPT PAG	0.9552	YES	NO	NO	NO

Multi-epitope Construct Assemblage and Physicochemical Functions

After thorough predictions and selection to create the multi-epitope sequence, only nine (9) epitopes {Linear Conformational B-cells (LCB) and HTL (*IL-4* and *IL-10* inducing)} emerged. They were linked with GSGSG, a flexible linker which made the final construct 162 amino acids long (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 16584.63 Da and a theoretical isoelectric point (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 hours, >20 hours, and >10 hours 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. So, 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 best appropriate model for further study since 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, which is the most appropriate, has a Z-score of -3.41 based on the Prosa-eb 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 (Fig. 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 Disulfide Engineering

Only one pair of amino acids, ALA 120 - GLY 123, was discovered to satisfy the criteria for the formation of disulfide bonds after residues were screened with DbD2 using the requirements of a χ^3 value of -87° or $+97^\circ$ and an energy value of 2.5 B-factor. These residues were then switched out for cysteine residues (Figure 9).

Molecular Docking & 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 - 1.000 (Table 3).

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

SOURCE	PIF	PIF	PIF	ENVIRONMENT	ALFALFA SPROUTS
GENES	<i>BcsC</i>	<i>SecY</i>	<i>FlhE</i>	<i>OmpA</i>	<i>FlhE</i>
DOCKING SCORES	-328.48	-369.53	-260.63	-282.06	-263.95
CONFERENCE SCORE	0.9726	0.9878	0.9014	0.9335	0.9071

In-Silico Cloning

The improved sequences exhibited a Codon Adaptation Index (CAI) value of 1.0 and a CG% (percentage of cytosine and guanine nucleotides) of 55.35%. In contrast, that of *Escherichia coli* (strain K12) was 50.73%. For the insilico 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 2683bp 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 septicemia, necrotising enterocolitis, and meningitis in neonates and infants. The xerotolerant pathogen uses a variety of genes to ensure its survival in dry conditions; therefore, it could be detected in stored powdered infant formula (PIF) even after 2.5years of storage (Srikumar *et al.*, 2019; Elkhawaga *et al.*, 2020). Currently, the detection methods used for *Cronobacter sakazakii* either have a longer turnaround time, require specialised 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 immunoinformatic 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; 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 & 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 needing a lysis buffer.

Continuous and discontinuous B-cell epitopes were predicted with three servers (IEDB, SVMTrip, 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 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 the B-cells' survival, growth and generation of antibodies, as supported by Bacharier and Geha (1952) and Moore et al. (2001). The predicted epitopes that passed all the necessary screenings were linked with the GSGSG linker to increase the construct folding and biological activities. 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 in 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 like 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 hours; in *E. coli*, the estimated half-life is > 10 hours in vivo and a half-life of 100 hours in mammalian reticulocytes in vitro. The GC% (guanine-cytosine content) and AI-value (codon adaptation index) are essential for successful protein expression in the prokaryotic host. Typically, the GC% falls within the 35-65% range, while the CAI-value ranges from 0.8 to 1. In this study, the JCAT online server was employed to enhance the GC% and CAI-value of the multi-epitope protein to 55.35% and 1.0, respectively, aiming 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 with the pET 28a(+) vector using the SnapGene software application for cloning purposes.

To precisely identify and cleave the DNA sequences in a controlled manner, restriction enzymes BstApl and TatI were employed. 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 the proper restriction enzyme sites.

SOPMA server analysis of the protein secondary structure shows the percentage of random coils, extended strands, beta (β) turn and alpha (α) helices to be 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 provided, but the best model was model 1, which had a better Ramo-favoured score (95.6) and GDT-HA score (0.8858). The Prosa-Web and MolProbity servers were utilised to validate the refining process's quality further. The Prosa-web results demonstrated a significant improvement, indicated by a Z-score of -3.41. Also, Ramachandran analysis by mol probity showed that the refined model had 95.6% and 97.5% of its amino acids in the Ramachandran plot's favoured region, which indicates that the refined model is of high quality. 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 lesser quality compared to the refined model. The result from DbD2 showed that the sequence was mutated at ALA120 and GLY 123 residue length 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, validations and disulphide engineering processes, was docked against the three antigenic genes whose source was from Powdered Infant Formula (PIF). It was examined that the docking score was < -200 . A similar technique was carried out for antigenic genes from other sources, such as the environment and alfalfa sprouts. However, their binding scores were less than that of the PIF isolates except for *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*.

Conclusion

Using bioinformatics and computational methods to design a detection kit for detecting *Cronobacter 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 *Cronobacter 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 lowering death rates. However, it is critical to recognise that further confirmation using in vitro and in vivo methods is required.

Acknowledgement

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

Author Contributions

Elijah K. Oladipo, Helen Onyeaka : Conceptualization, Supervision, Editing, Elijah K. Oladipo, Temitope A. Oyedepo: Original draft preparation, Methodology, Olukayode I. Obisanya , Victoria O. Owoeye, Oyinlola G. Shittu : Writing, Reviewing, Mautin G. Adamitonde, Emmanuel C. Ikwuka, Solomon O Ojewale: Methodology, Data Analysis, Amudatu A. Adedokun ,Adewale J. Tijani, Feranmi A. Adedokun: Software, Investigation.

Accepted Manuscript

References

- Aly, M. A., Reimhult, E., Kneifel, W., & Domig, K. J. (2019). Characterization of biofilm formation by *Cronobacter* spp. Isolates of different food origin under model conditions. *Journal of Food Protection*, 82(1), 65-77
- Bacharier, L. B., & Geha, R. S. (2000). Molecular mechanisms of IgE regulation. *Journal of Allergy and Clinical Immunology*, 105(2), S547-S558.
- Cai, X. Q., Yu, H. Q., Ruan, Z. X., Yang, L. L., Bai, J. S., Qiu, D. Y., ... & Zhu, X. Q. (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., Cheetham, P., Loc-Carrillo, C., Fayet, O., ... & Forsythe, S. J. (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 2.0: a web-based tool for disulfide engineering in proteins. *BMC bioinformatics*, 14(1), 1-7.
- 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.
- Ebrahimi, M., Seyyedtabaei, S. J., Ranjbar, M. M., Tahvildar-biderouni, F., & Javadi Mamaghani, A. (2020). Designing and modeling of multi-epitope proteins for diagnosis of *Toxocara canis* infection. *International Journal of Peptide Research and Therapeutics*, 26, 1371-1380.

Elkhawaga, A. A., Hetta, H. F., Osman, N. S., Hosni, A., & El-Mokhtar, M. A. (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., Forsythe, S. J., Chai, Y., Li, R., ... & Jiang, Y. (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., Dong, G., Chang, Y., Meng, Z., ... & Yang, W. (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.

Ferrè, F., & Clote, P. (2005). DiANNA: a web server for disulfide connectivity prediction. *Nucleic acids research*, 33(suppl_2), W230-W232.

Gao, X., Dong, X., Li, X., Liu, Z., & Liu, H. (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. *Bioinformatics*, 11(6), 681-684.

Grote, A., Hiller, K., Scheer, M., Münch, R., Nörtemann, B., Hempel, D. C., & Jahn, D. (2005). JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic acids research*, 33(suppl_2), W526-W531.

Hariri, S., Joseph, S., & Forsythe, S. J. (2013). *Cronobacter sakazakii* ST4 strains and neonatal meningitis, United States. *Emerging Infectious Diseases*, 19(1), 175.

Hartmann, I., Carranza, P., Lehner, A., Stephan, R., Eberl, L., & Riedel, K. (2010). Genes involved in *Cronobacter sakazakii* biofilm formation. *Applied and Environmental Microbiology*, 76(7), 2251-2261.

Healy, B., Cooney, S., O'Brien, S., Iversen, C., Whyte, P., Nally, J., ... & Fanning, S. (2010). *Cronobacter* (*Enterobacter sakazakii*): an opportunistic foodborne pathogen. *Foodborne Pathogens and Disease*, 7(4), 339-350.

Hebditch, M., Carballo-Amador, M. A., Charonis, S., Curtis, R., & Warwicker, J. (2017). Protein-Sol: a web tool for predicting protein solubility from sequence. *Bioinformatics*, 33(19), 3098-3100.

Iversen, C., Lehner, A., Mullane, N., Bidlas, E., Cleenwerck, I., Marugg, J., ... & Joosten, H. (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.

Li, C., Sun, X., Niu, B., Jiang, Y., Yang, J., & Chen, Q. (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., Jensen, S. O., van Hal, S. J., & Davis, R. (2018). Cronobacter sakazakii infection from expressed breast milk, Australia. *Emerging Infectious Diseases*, 24(2), 393.

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S. & Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nature Methods*, 19(6), 679-682.

Oladipo, E. K., Jimah, E. M., Irewolede, B. A., Folakanmi, E. O., Olubodun, O. A., Adediran, D. A., ... & Adelusi, T. I. (2022). Immunoinformatics design of multi-epitope peptide for the diagnosis of Schistosoma haematobium infection. *Journal of Biomolecular Structure and Dynamics*, 1-8.

Oladipo, E. K., Oluwasegun, J. A., Oladunni, T. D., Owoeye, V. O., Obisanya, O. I., Olasinde, O. T., & Onyeaka, H. (2023). Immunoinformatics aided design of a peptide- based kit for detecting Escherichia coli O157: H7 from food sources. *Journal of Food Safety*, e13073.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., ... & Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualisation for researchers, educators, and developers. *Protein Science*, 30(1), 70-82.

Ponomarenko, J., Bui, H. H., Li, W., Füsseder, N., Bourne, P. E., Sette, A., & Peters, B. (2008). ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics*, 9, 1-8.

Remmert, M., Biegert, A., Hauser, A., & Söding, J. (2012). HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nature Methods*, 9(2), 173-175.

Rose, R., Golosova, O., Sukhomlinov, D., Tiunov, A., & Prosperi, M. (2019). Flexible design of multiple metagenomics classification pipelines with UGENE. *Bioinformatics*, 35(11), 1963-1965.

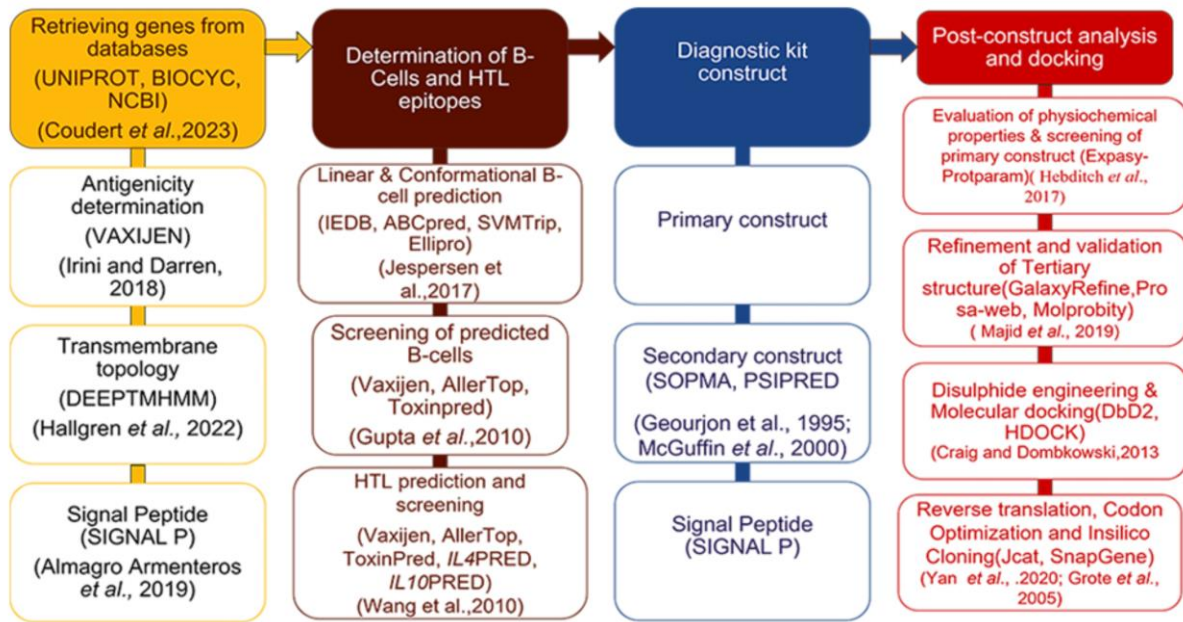
Safavi, A., Kefayat, A., Sotoodehnejadnematalahi, F., Salehi, M., & Modarressi, M. H. (2019). Production, purification, and in vivo evaluation of a novel multiepitope peptide vaccine consisted of immunodominant epitopes of SYCP1 and ACRBP antigens as a prophylactic melanoma vaccine. *International immunopharmacology*, 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.

Schoch, C. L., Ciuffo, S., Domrachev, M., Hotton, C. L., Kannan, S., Khovanskaya, R., Leipe, D., Mcveigh, R., O'Neill, K., Robbertse, B., Sharma, S., Soussov, V., Sullivan, J. P., Sun, L., Turner, S., & Karsch-Mizrachi, I. (2020). NCBI Taxonomy: a comprehensive update on curation, resources and tools. Database: *The Journal of Biological Databases and Curation*. <https://doi.org/10.1093/database/baaa062>

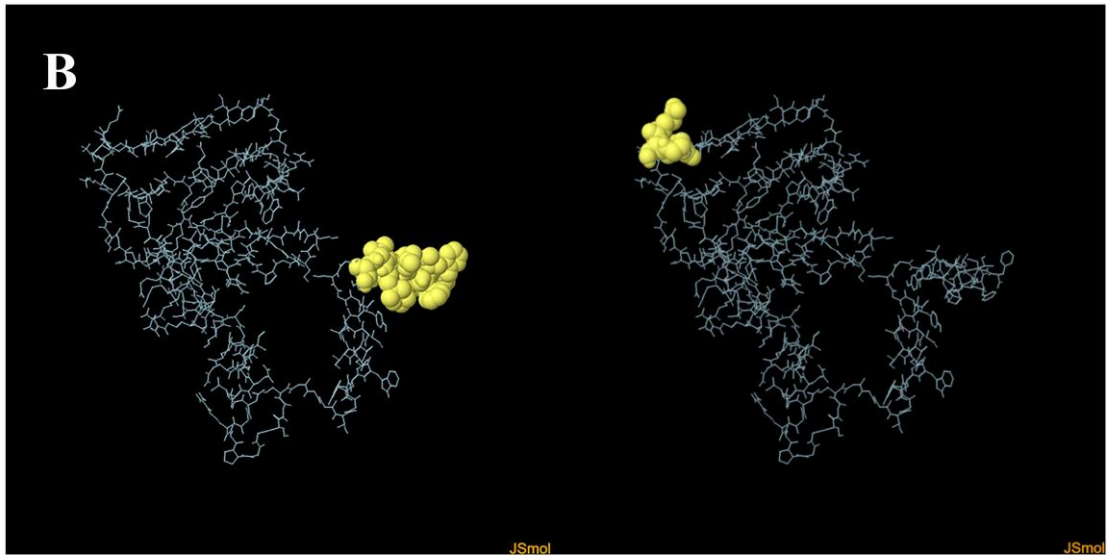
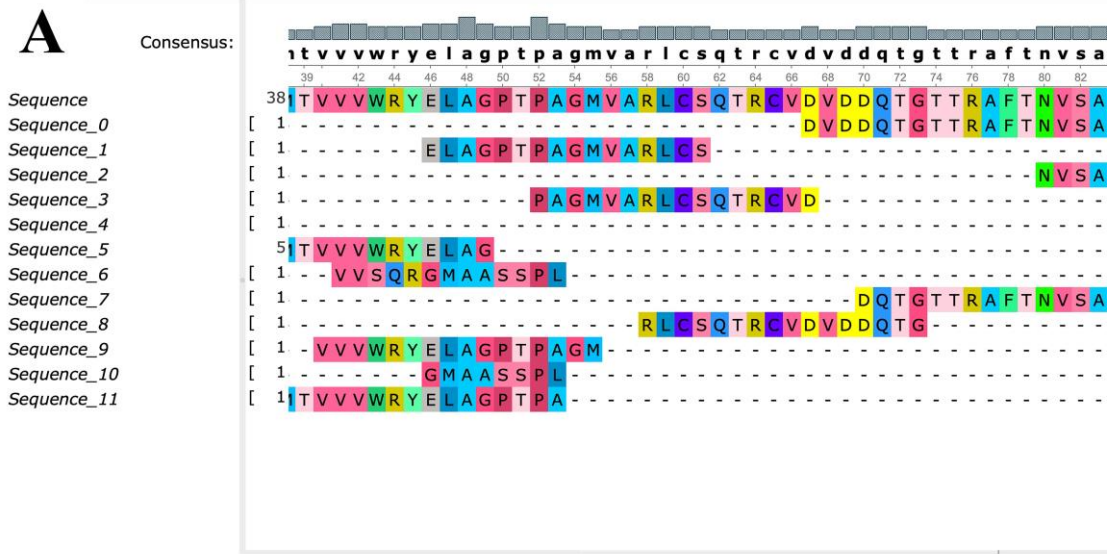
- Shams, M., Nourmohammadi, H., Asghari, A., Basati, G., Majidiani, H., Naserifar, R., & Irannejad, H. (2021). Construction of a multi-epitope protein for human *Toxocara canis* detection: *Immunoinformatics approach multi-epitope construct for T. canis serodiagnosis. Informatics in Medicine Unlocked*, 26, 100732.
- Srikumar, S., Cao, Y., Yan, Q., Van Hoorde, K., Nguyen, S., Cooney, S., ... & Fanning, S. (2019). RNA sequencing-based transcriptional overview of xerotolerance in *Cronobacter sakazakii* SP291. *Applied and Environmental Microbiology*, 85(3), e01993-18.
- 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(suppl_2), W407-W410.
- Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., ... & Richardson, D. C. (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., & Zhang, C. (2012). SVMTriP: a method to predict antigenic epitopes using a support vector machine to integrate tri-peptide similarity and propensity.
- Yao, M., Guo, X., Wu, X., Bai, Q., Sun, M., & Yin, D. (2022). Evaluation of the Combined Use of Major Outer Membrane Proteins in the Serodiagnosis of Brucellosis. *Infection and Drug Resistance*, 4093-4100.
- Zheng, D., Liang, S., & Zhang, C. (2022). B-Cell Epitope Predictions Using Computational Methods. *In Computer-Aided Antibody Design* (pp. 239-254). New York, NY: Springer US.

Figure 1



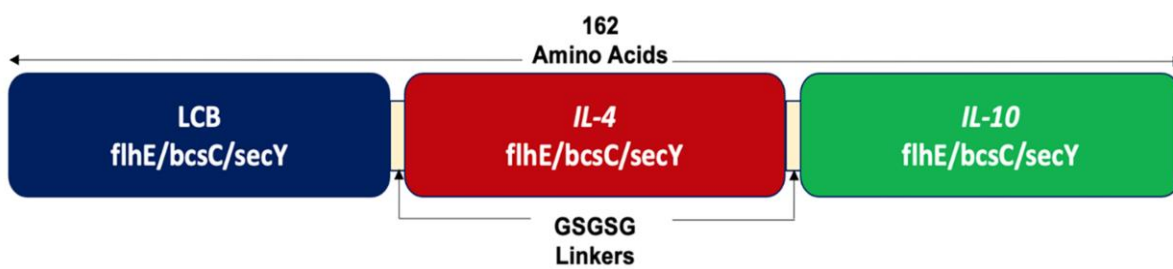
Accepted Manuscript

Figure 3



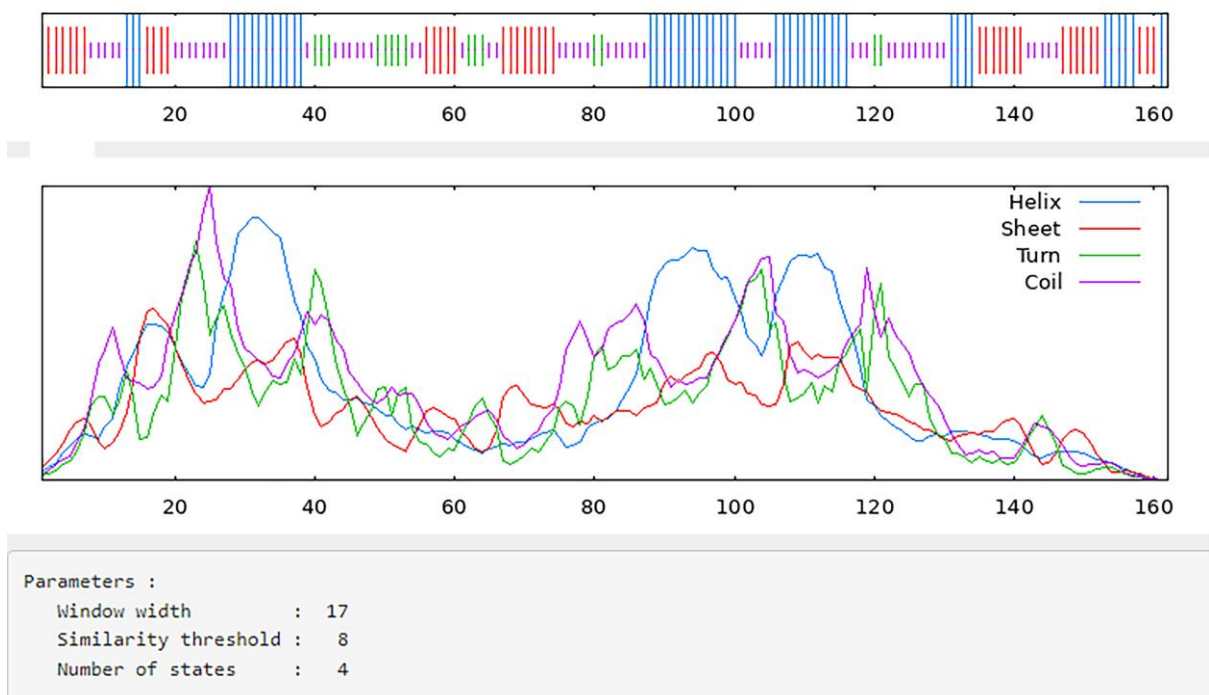
ACCEPT

Figure 4



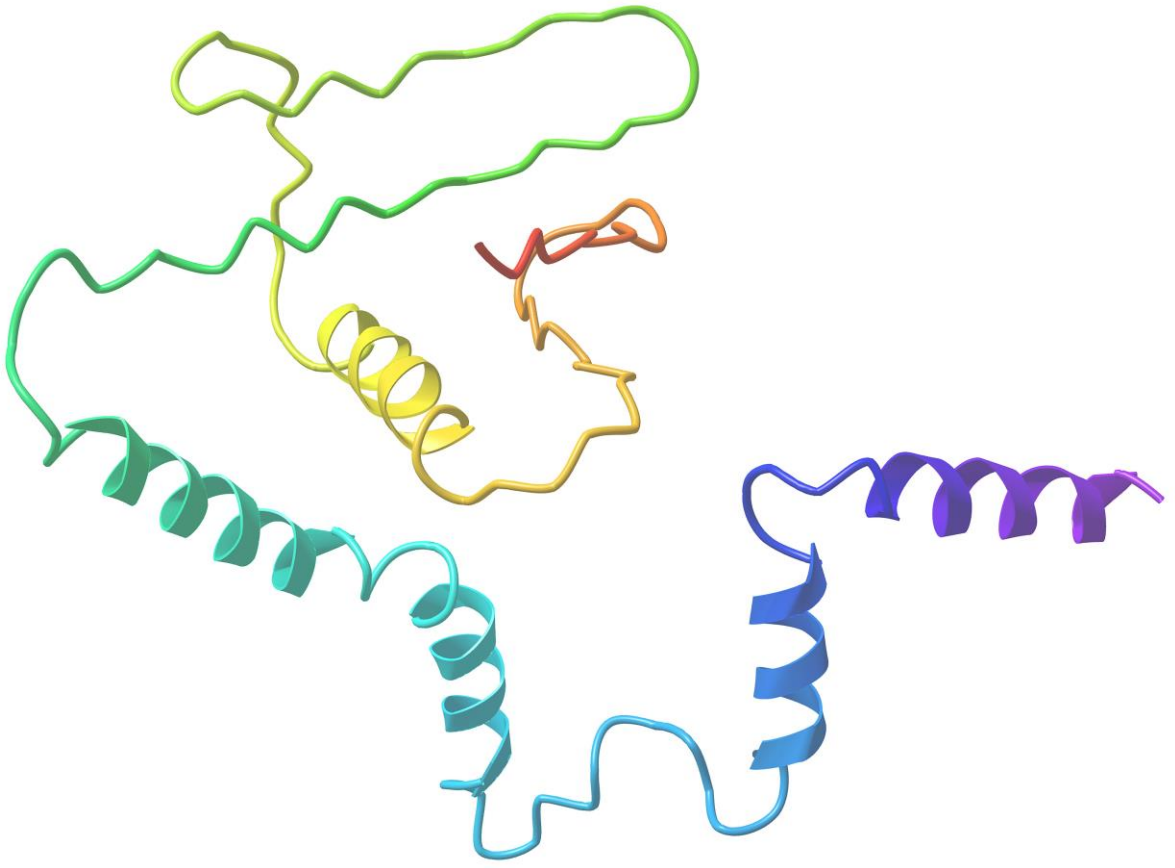
Accepted Manuscript

Figure 5



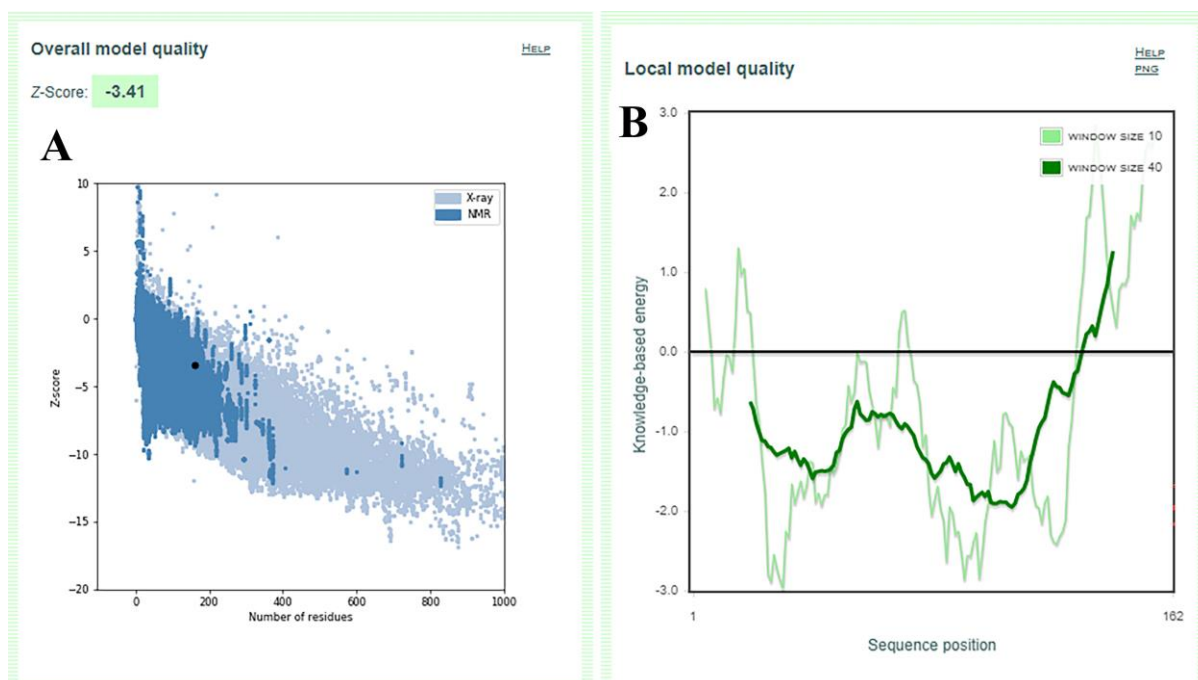
Accepted Manuscript

Figure 6



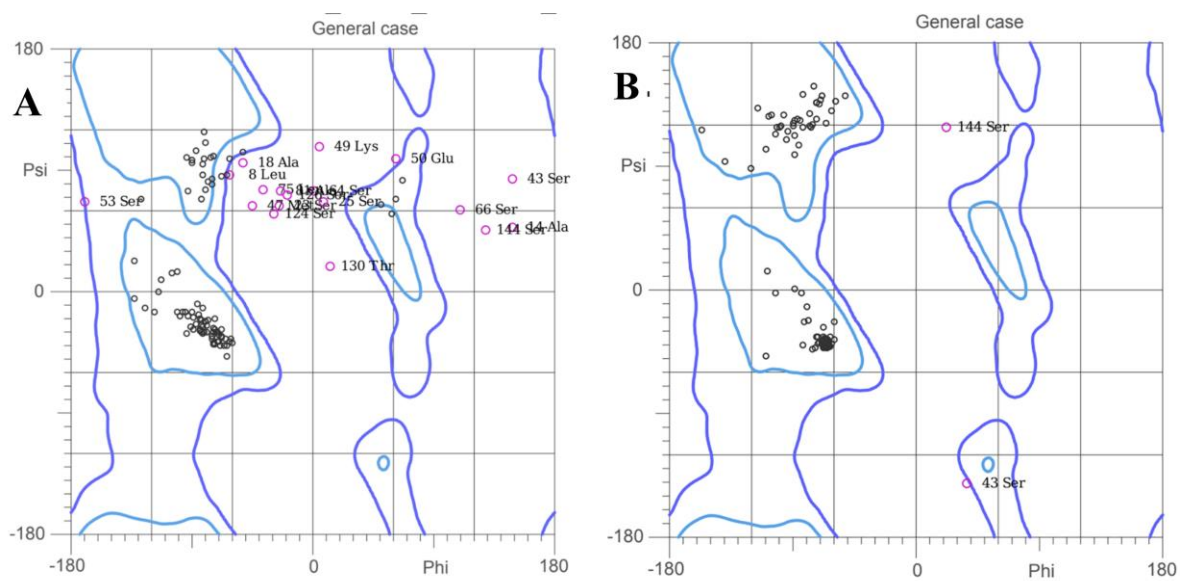
Accepted

Figure 7



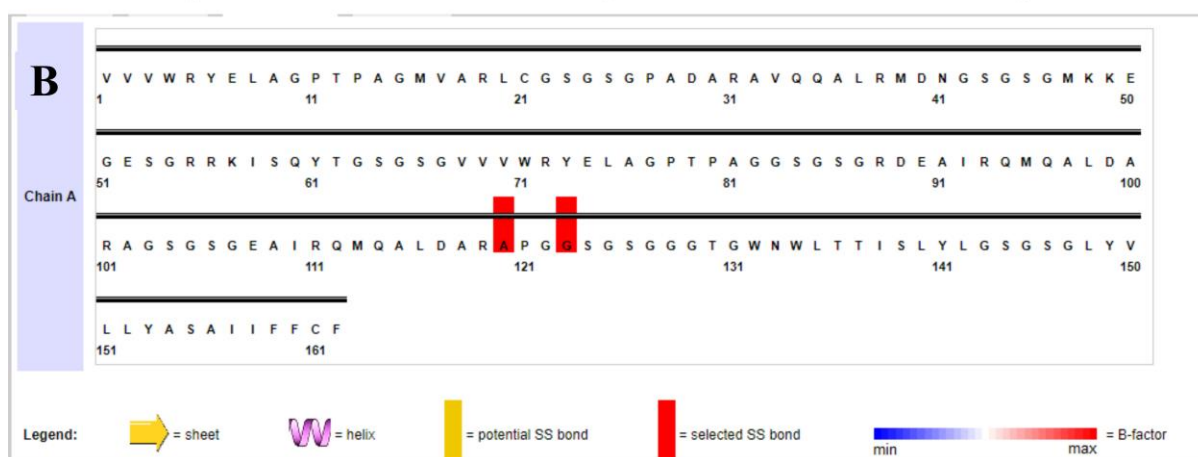
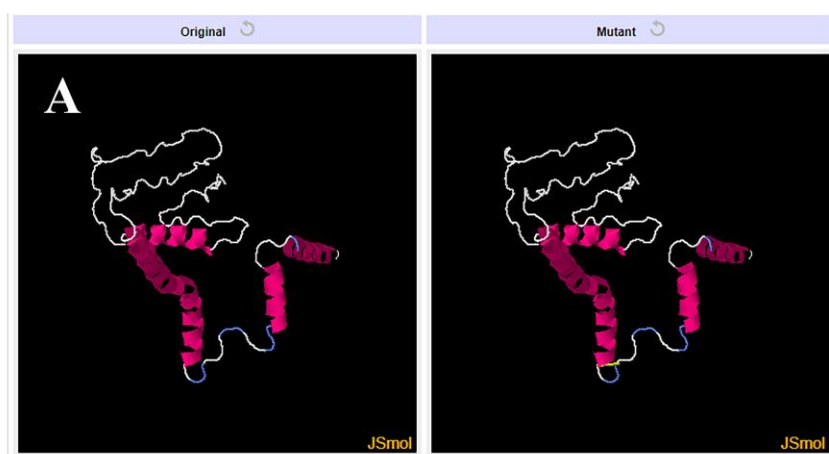
Accepted Manuscript

Figure 8



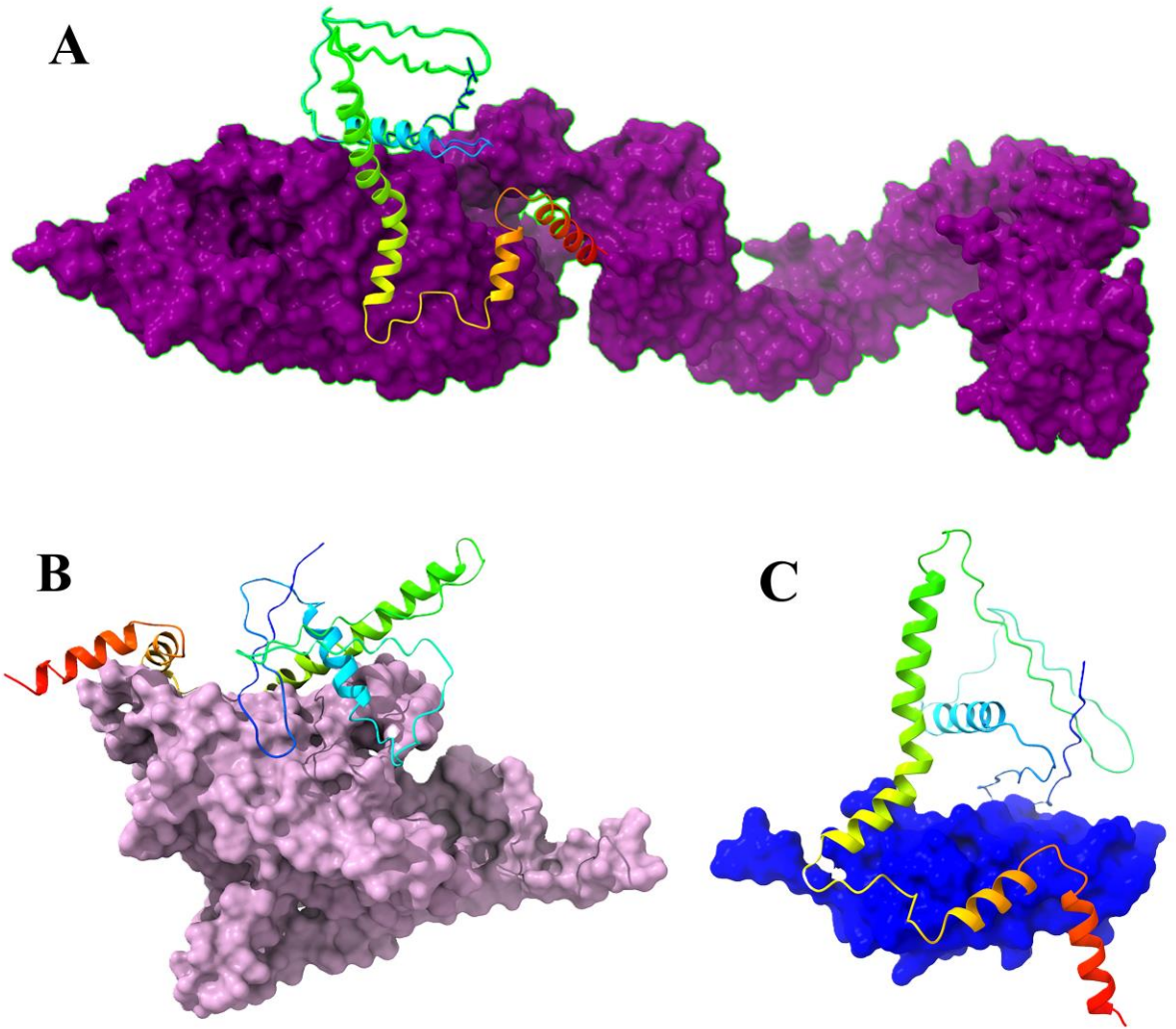
Accepted Manuscript

Figure 9



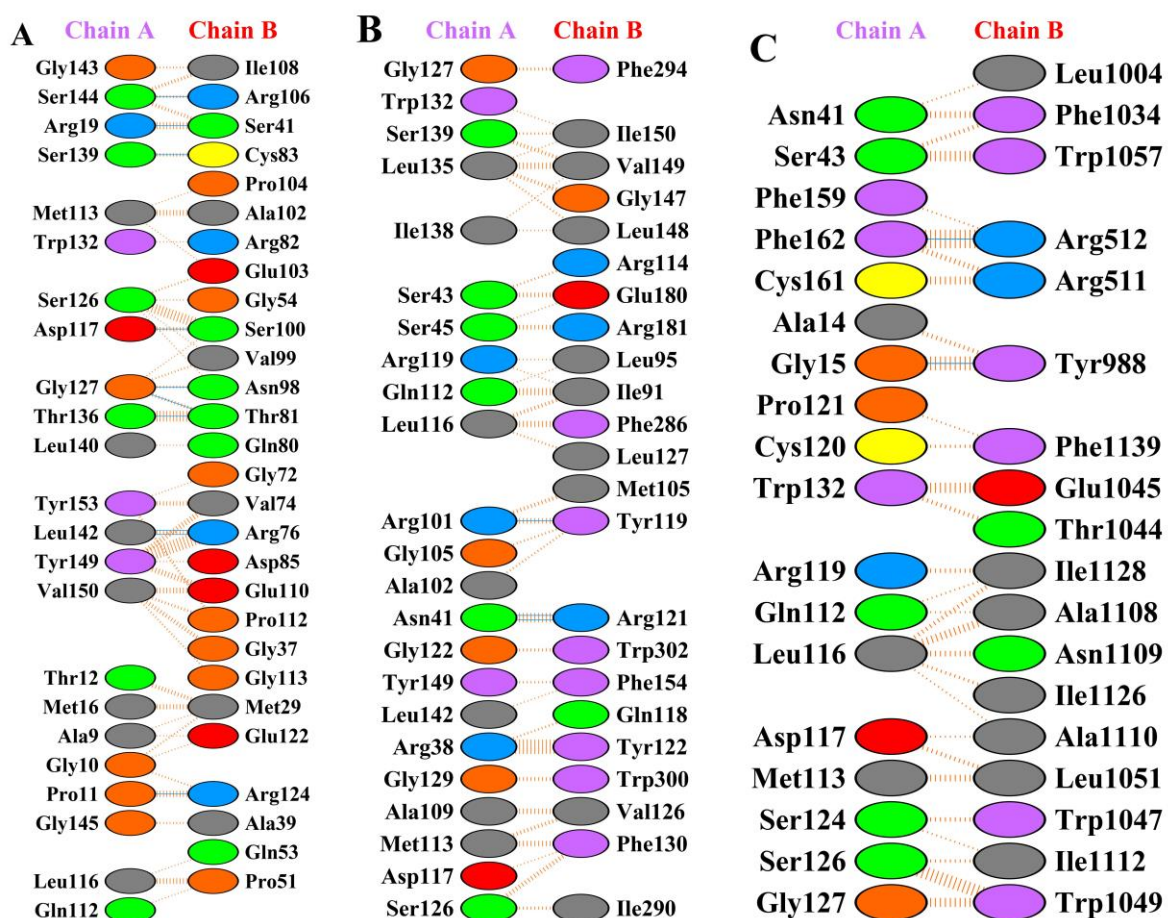
Accepted

Figure 10



Accepted

Figure 11

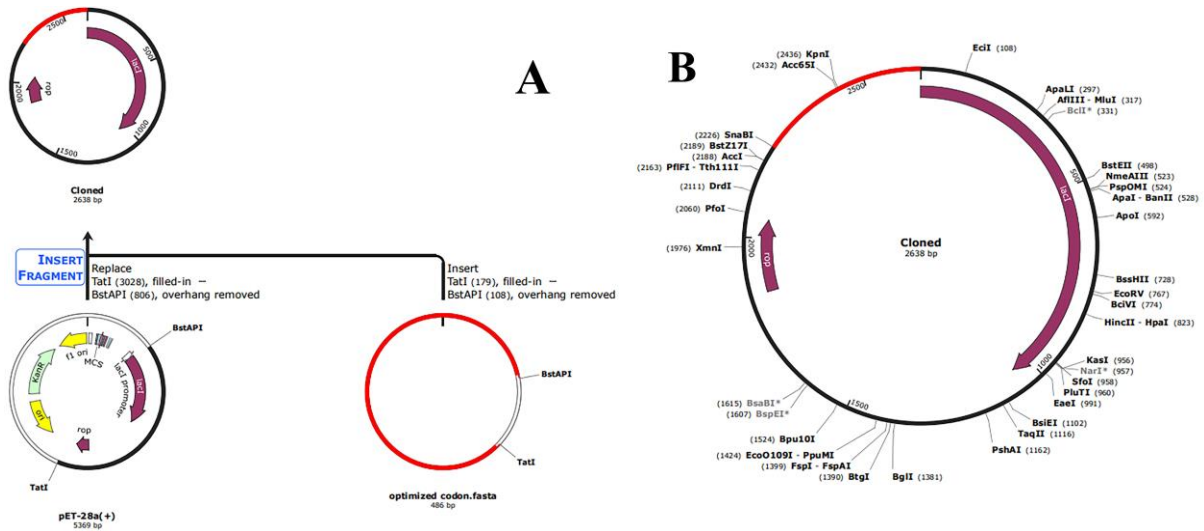


D **Key:** — Salt bridges — Disulphide bonds — Hydrogen bonds — Non-bonded contacts

The number of H-bond lines between any two residues indicates the number of potential hydrogen bonds between them. For non-bonded contacts, which can be plentiful, the width of the striped line is proportional to the number of atomic contacts.

Residue colours: Positive (H,K,R); negative (D,E); S,T,N,Q = neutral; A,V,L,I,M = aliphatic; F,Y,W = aromatic; P,G = Pro&Gly; C = cysteine.

Figure 12



Accepted Manuscript