UNIVERSITYᴼᶠ BIRMINGHAM

University of Birmingham
Research at Birmingham

Immunoinformatics aided design of a peptide-based kit for detecting Escherichia coli O157: H7 from food sources

Oladipo, Elijah Kolawole; Oluwasegun, Jerry Ayobami; Oladunni, Tolu David; Owoeye, Victoria Oluwakemi; Obisanya, Olukayode Ifeanyi; Olasinde, Oluwatosin Tumininu; Onyeaka, Helen

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

ORIGINAL ARTICLE

Immunoinformatics aided design of a peptide-based kit for detecting *Escherichia coli* O157:H7 from food sources

Elijah Kolawole Oladipo[1,2] | Jerry Ayobami Oluwasegun[2,3] |
Tolu David Oladunni[2] | Victoria Oluwakemi Owoeye[4] |
Olukayode Ifeanyi Obisanya[2,5] | Oluwatosin Tumininu Olasinde[6] |
Helen Onyeaka[7]

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

[2]Genomics unit, Helix Biogen Institute, Ogbomoso, Oyo, Nigeria

[3]Department of Physiology, Ladoke Akintola University of Technology, Ogbomoso, Oyo, Nigeria

[4]Department of Microbiology, Obafemi Awolowo University, Ile-Ife, Osun, Nigeria

[5]Department of Medical Laboratory Science, College of Medicine, University of Lagos, Lagos, Nigeria

[6]Department of Biochemistry, Ladoke Akintola University of Technology, Ogbomoso, Oyo, Nigeria

[7]School of Chemical Engineering, University of Birmingham, Birmingham, UK

Correspondence

Elijah Kolawole Oladipo, Department of Microbiology, Laboratory of Molecular Biology, Immunology and Bioinformatics; Genomics Unit, Helix Biogen Institute, Ogbomoso, Oyo State, Nigeria.
Email: koladipo2k3@yahoo.co.uk

Helen Onyeaka, School of Chemical Engineering, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK.
Email: h.onyeaka@bham.ac.uk

Abstract

Food and water-borne enterohemorrhagic *Escherichia coli* (EHEC) serotype O157:H7 is a zoonotic bacterium that causes gastroenteritis and other human diseases. It has also been linked to chronic foodborne diseases with high mortality rates worldwide, particularly in children. Hence, this study was carried out to designed a peptide base kit for quick detection of *E. coli* in food. A peptide-based rapid detection kit was designed using an immunoinformatic technique and some antigenic target genes (*stx1A*, *stx2B*, *escC*, *fliC*, and *eae*). The antigenic gene sequences retrieved were screened for antigenicity, transmembrane topology, B-cells and helper T-cells. Selected epitopes were joined with appropriate linkers to form a chimeric protein which consists of five B-cell epitopes, five interleukin-4 (*IL-4*) inducer epitopes and five interleukin-10 (*IL-10*) inducer epitopes. The improved and optimized chimeric protein sequence was cloned in-silico in a suitable expression host, *E. coli*-strain K12. The designed peptide refined and validated tertiary structure was molecularly docked with the tertiary structures of each antigenic target gene. The physicochemical properties of the chimeric protein showed that the construct has an amino acid length of 295 amino acids, a molecular weight of 29.876 kiloDalton (kDa), an aliphatic index of 75.05 and an instability index of 14.82 which confers stability. The construct was hydrophilic with a GRAVY value of -0.261 and had a considerable half-life of 4.4 h (mammalian reticulocytes, in vitro), >20 h (yeast, in vivo) and >10 h (*E. coli*, in vivo). Conclusively, the final construct has successfully met the design requirements for the development of a lateral flow kit, which has the potential to provide fast and efficient detection of *E. coli* O157:H7. However, it is the additional validation through the vitro and in vivo techniques needed to confirm that this designed peptide based test kit.

1 | INTRODUCTION

Escherichia coli (*E. coli*) is one of the most extensively researched microbes. It belongs to the Enterobacteriaceae family, which includes some of the most significant intestinal pathogens, including *Yersinia*, *Shigella*, and *Salmonella*. Although most *E. coli* are commensal and non-pathogenic, naturally occurring in the intestinal tracts of various animals and humans, specific clusters of this microbe have been linked to a range of intestinal and extraintestinal diseases. These illnesses can result in severe complications, such as life-threatening diarrhea and serious disabilities in animals and humans (Kaper et al., 2004). Shiga toxin-producing *E. coli* (STEC), which includes enterohemorrhagic *E. coli* (EHEC), is a broad classification distinguished by the production of Shiga toxins (Stx). There are approximately 300 serotypes of STEC. They are zoonotic extracellular pathogens that occasionally cause diarrhea outbreaks, hemolytic uremic syndrome and hemorrhagic colitis, common in developed countries, but not all serotypes have been implicated in illness (Brooks et al., 2005; Nguyen & Sperandio, 2012). Specific serotypes of pathogenic *E. coli*, such as STEC O157:H7, O145:H28, and O103:H21, have been specifically associated with some particular pathotypes of the bacterium. The serotype most frequently associated with sporadic cases and diarrhea outbreaks, O157:H7 (Karmali et al., 2010), has been linked to over 63,000 illnesses per year, and more than 2100 hospitalizations and deaths in the United States (Fatima & Aziz, 2023). It has also been repeatedly linked to hospitalizations and deaths, particularly among young children and the elderly (Marks et al., 2013). These outbreaks worldwide are attributed to the consumption of contaminated food, including but not limited to undercooked beef, roasted beef, fruit/vegetable salad, cooked meats, venison meat and jerky, radish sprouts, salami, cheese, ice cream bars, raw milk, pasteurized milk, yoghurt, lettuce, spinach, unpasteurized apple cider/ juice, cantaloupe, potatoes, alfalfa sprouts and cake, as the primary reservoir of EHEC are ruminant animals especially cattle (Kaper et al., 2004).

Studies and investigations on EHEC O157:H7 reveal that its pathogenicity is associated with the utilization of some antigenic genes like EscC (T3SS-type III secretion system), Intimin, Shiga-like toxin (Stx) and fliC (flagellin coding). EscC (T3SS-type III secretion system) is a channel for passing effector proteins into cells, ultimately changing the host cell cytoskeleton and causing the death of the host cell in some cases (Wong et al., 2011). The bacterial adhesion molecule intimin (a mature protein of 101 kDa) is linked to close contact between bacteria (translocated intimin receptor [Tir-Int] interaction) (Batchelor et al., 2000). Another essential virulence component of EHEC O157:H7 is the Shiga-like toxin (Stx) protein known as verotoxin due to its ability to kill Vero cells (Bunger et al., 2015). The Stx belong to the AB5 toxins family, consisting of Stx1 and Stx2 subgroups (Bunger et al., 2015), which are sometimes responsible for hemorrhagic colitis, severe bloody diarrhea, and HUS syndrome (Mellmann et al., 2008). Studies also reveal that the lipopolysaccharide (LPS) present in the bacteria's outer membrane contains an O-side chain which exhibits enormous antigenic variability and determines the serological specificity. In a study by Mahajan et al. (2009) and Rani et al. (2021), the fliC gene, which encodes flagellin (the organelle responsible for motility), has been shown to participate in adherence of the pathogen to the epithelial cells and was used for the detection of *E. coli* O157:H7 in commercial foods.

Considering the massive public health consequences EHEC O157:H7 has caused regarding food hygiene and safety in the food industry, where microbiological safety is a problem in particular, several methods have been implemented to detect the organism in food and water rapidly. Conventional methods that have been in use include plating and culturing, use of biochemical tests—modified agar methods (Deisingh & Thompson, 2004), Molecular assays like polymerase chain reaction (PCR) (Loge et al., 2002), pulse field gel electrophoresis (PFGE) (Ribot et al., 2006), and enzyme-linked immunosorbent assay (ELISA) (Gehring & Tu, 2005). Even though these methods were largely accurate, there are some shortcomings: low turnaround time, expensive and labor-intensive, requires sophisticated devices with highly trained personnel (Oladipo et al., 2020). Unlike diagnostic kits used in recent outbreaks as recommended by the WHO, these rendered the approaches inadequate for real-time monitoring of any disease outbreak. (Rani et al., 2021; Zhao et al., 2018). The main purpose of this study is to identify probable genes of *E. coli* from food sources that can be used in the design of a peptide-based detection kit for EHEC O157:H7.

2 | MATERIALS AND METHODS

2.1 | Protein sequence retrieval

The associated genes of Stx 2B (Shiga toxin 2), Stx 1A (Shiga toxin 1), FliC (flagellar filament structural protein), EscC (T3SS—type III secretion system structure protein), and Intimin proteins are *stx2B*, *stx1A*, *fliC*, *escC*, and *eae* respectively. These genes are essential in the toxicity and attachment of *E. coli* O157:H7, which have been incriminated in many foodborne diseases outbreaks of diarrhea hemorrhagic colitis (García-Angulo et al., 2014; Najafabadi et al., 2021). Their sequences were retrieved from Ecocyc, an online *E. coli* database linked to NCBI (National Center for Biotechnology Information) and UniProtKB with their accession numbers (Appendix S1) (Figure 1).

2.2 | Antigenicity of the retrieved sequences

Though the selected genes have been identified from literature as antigenic, their antigenicity was predicted using Vaxijen, an alignment-free approach (Doytchinova & Flower, 2007). The "bacteria" model on Vaxijen was used, and genes with antigenicity score ≥0.5 were selected. DeepTMHMM was used to predict the membrane topology of the antigenic genes (Hallgren et al., 2022). DeepTMHMM
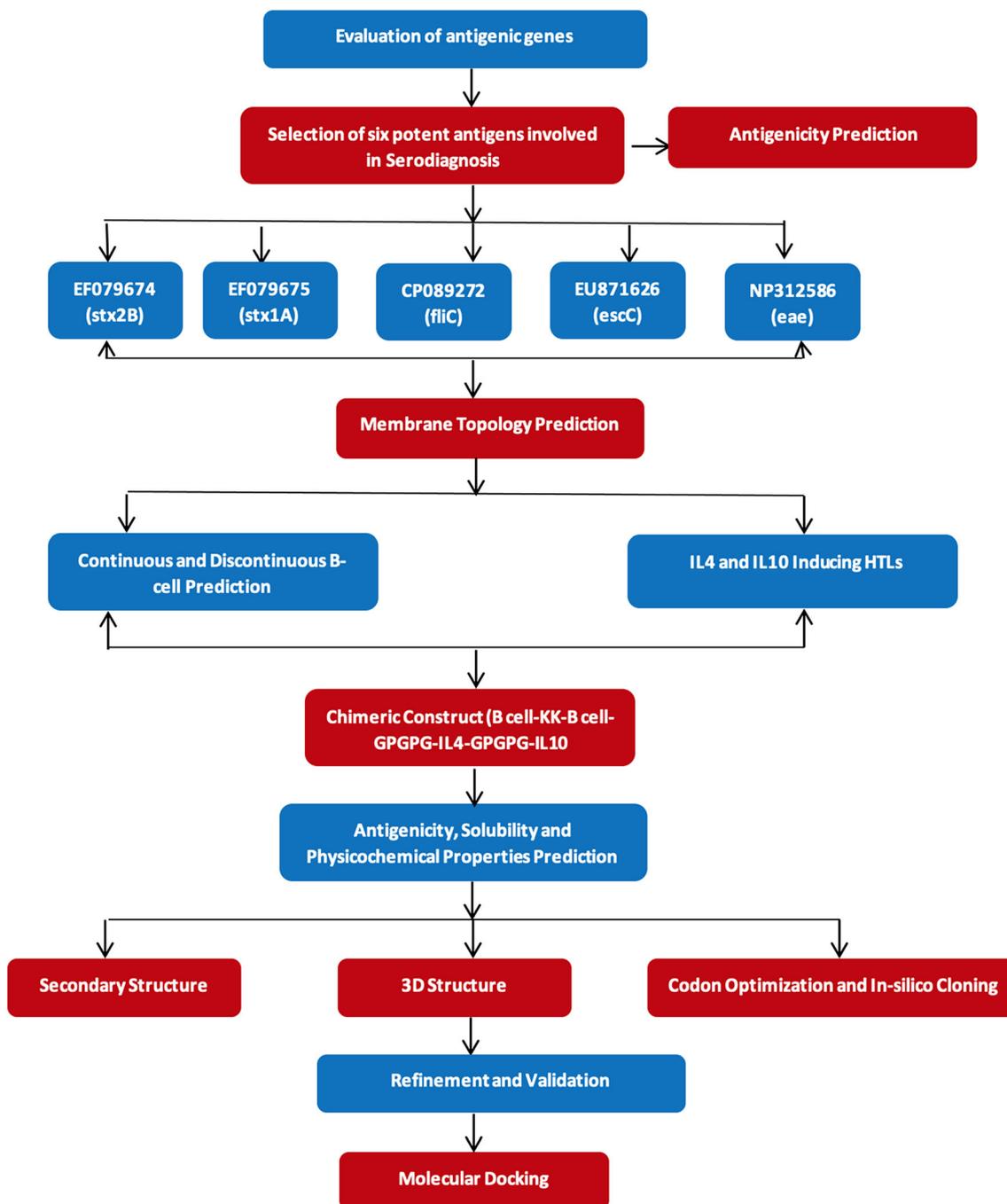
FIGURE 1 Flowchart showing the steps used in the design of the peptide-based kit.

classifies membrane topology into a signal peptide, inside, transmembrane, and outside.

2.3 | Continuous and discontinuous B cell epitopes

Bepipred 2.0, IEDB and SVMTrip web servers were employed to predict continuous B-cell epitopes. These tools use algorithms like Protein Language Models (Clifford et al., 2022) and Support Vector Machine (Yao et al., 2020). The continuous B cell epitopes across the three servers were aligned with Unipro UGENE using the MAFFT

algorithm (Weidmann et al., 2018) to get the consensus epitopes across the server. Also, the ElliPro server predicted discontinuous B-cell epitopes with the tertiary structure of the antigenic genes (Ponomarenko et al., 2008).

2.4 | Helper T lymphocyte epitopes

The epitopes database server NETMHC-II 4.0 (Reynisson et al., 2020) predict helper T lymphocyte (HTL) epitopes. This server predominantly presents peptides from extracellular proteins and informs if the

sequence is a strong or a weak MHC binder based on a 1% and 5% rank score, respectively. NETMHC-II 4.0 predicts binding affinities of the peptide to MHC-II molecules covering the selected species/loci HighQ-DRB.

2.5 | Antigenicity, allergenicity and toxicity of IL4 and IL10 inducers

IL4Pred (Dhanda et al., 2013) and IL10 Pred (Nagpal et al., 2017) servers were used to determine the selected HTL epitopes' ability to induce the production of interleukin 4 and interleukin 10. The IL4 and IL10-inducing epitope antigenicity were predicted with Vaxijen v2.0, allergenicity with AllerTOP v. 2.0 (Dimitrov et al., 2014), and toxicity using Toxinpred (Gupta et al., 2015).

2.6 | Chimeric protein construct

The chimeric protein construct of the diagnostic kit was constructed using the predicted B-cell and HTL epitopes. The selected epitopes had the most antigenicity score, non-allergenic, non-toxic and are able to induce IL4 and IL10. The HTL and B cell epitopes were linked using GPGPG and KK, respectively (Kalita et al., 2019).

2.7 | Antigenicity, solubility and physicochemical properties prediction of the chimeric protein construct

Antigenicity and solubility prediction of the primary construct was executed with Vaxijen v2.0 and Protein-Sol (Hebditch et al., 2017), respectively. The physicochemical properties (the number of amino acids, molecular weight, theoretical pI, instability index, aliphatic index, atomic composition, estimated half-life and grand average of hydropathicity [GRAVY]) of the chimeric protein construct were computed using the Expasy Protparam web server (Gasteiger et al., 2003; Hennebert et al., 2015).

2.8 | Codon optimization and in-silico cloning

The amino acid sequence of the chimeric construct was reverse-translated into nucleotide sequence, after which Java Codon Adaptation Tool (JCAT) online server was used to compute the optimized codon (Grote et al., 2005). Codon optimization was carried out by adapting the chimeric protein construct into a specific expression host (*E. coli* K12) to provide an improved DNA with a good codon adaptation index (CAI) and percentage of GC content suitable for the expression host (Oladipo et al., 2022). SnapGene software was employed in cloning the optimized translated chimeric construct in pET-28a (+), a suitable expression vector (Sarker et al., 2019).

2.9 | Secondary and tertiary structure prediction

SOPMA (self-optimized prediction method with alignment), an online tool, was used to predict the secondary structure of the chimeric construct. It predicts beta sheets, alpha helices and coils from the submitted amino acid sequence (Geourjon & Deleage, 1995). Subsequently, the chimeric construct tertiary (3D) structure was modeled using Alphafold 2.0 server, which predicts the 3D atomic coordinates of folded protein structures at a global distance test total score (GDT-TS), with a median value of 92.4% (Mirdita et al., 2022).

2.10 | Tertiary structure validation and refinement

The best-adapted model suggested by Alphafold 2.0 was further refined and improved using GalaxyRefine (Heo et al., 2013). GalaxyRefine refines protein tertiary structure by rebuilding side chains before performing side-chain repacking and subsequent structural relaxation through molecular dynamics simulation. The 3D structure was validated with ProSA-web, which exploits interactive web-based applications to display energy and values plots that highlight possible errors in the protein structures (Wiederstein & Sippl, 2007). PDBSum was used to further validate the structure by generating a Ramachandran plot, which shows the phi(φ)–psi(ψ) torsion angles for all amino acid residues (Laskowski & Thornton, 2021).

2.11 | Molecular docking

The final refined and validated tertiary structure(ligand) was docked with the tertiary structures of the five genes—receptor (*stx1A*, *stx2B*, *escC*, *fliC*, and *eae*), used in epitopes prediction respectively to evaluate the binding affinity and interactions between them by using an online docking tool, HDOCK (Yan et al., 2020).

3 | RESULTS

3.1 | Retrieved protein sequences

The protein sequences retrieved from NCBI were *stx2B* (accession no. EF079674.1), *stx1A* (accession no. EF079675.1), *fliC* (accession no. CP089272.1), *escC* (accession no. EU871626.1), and *eae* (accession no. NP312586.1) (Table 1) in FASTA format.

3.2 | Antigenicity and membrane topology prediction

All the protein sequences used were antigenic with an antigenicity score ≥0.5 and had the whole or portions of their sequences located outside shown in Table S1 and Figure S1.

TABLE 1 Details, antigenicity and transmembrane topology of *E. coli* O157:H7 proteins.

No.	Accession number	Protein	Amino acids	Antigenic score (Vaxijen)	Transmembrane topology (DeepTMHMM)
1	EF079674.1	Stx 2B (Shiga toxin 2)	89aa	0.8063	Outside
2	EF079675.1	Stx 1A (Shiga toxin 1)	315aa	0.5137	Outside
3	CP089272.1	FliC (flagellar filament structural)	585aa	0.7694	Outside
4	EU871626.1	EscC (T3SS—type III secretion system structure)	508aa	0.5219	Inside and outside
6	NP312586.1	Eae(Intimin)	934aa	0.6784	Inside and outside

FIGURE 2 Alignment of predicted linear B-cell epitopes with MAFFT algorithm on Unipro UGENE showing consensus.
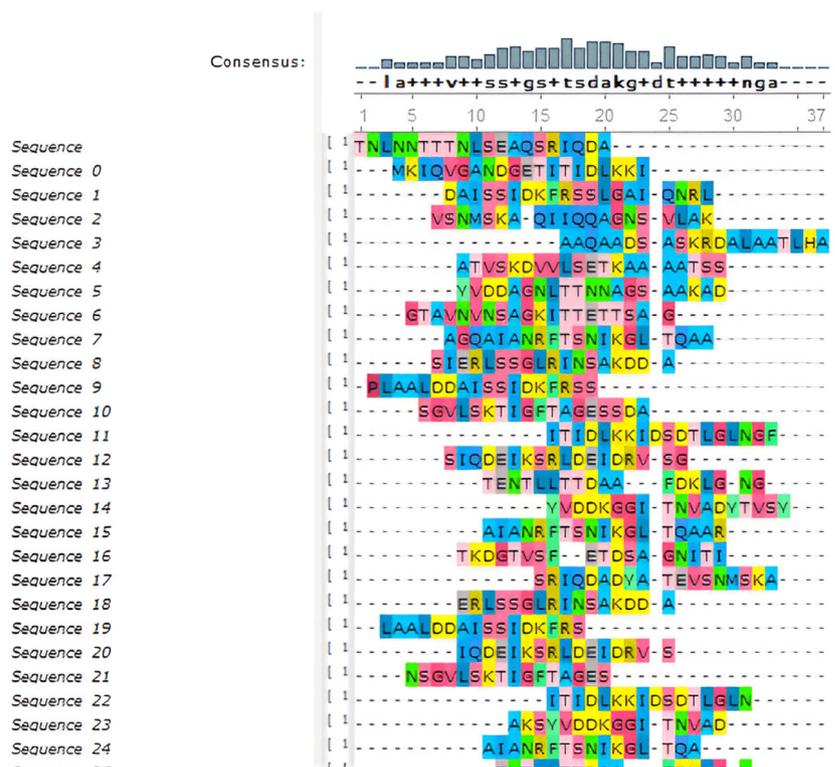
TABLE 2 Showing selected predicted IL4 and IL10 inducing HTL epitopes, their antigenicity score, allergenicity and toxicity.

HTL epitopes (*NETMHCII*)	IL-4 inducer (*IL4Pred*)	IL-10 inducer (*IL10Pred*)	Antigenicity score (*Vaxijen*)	Allergenicity (*AllerTOP*)	Toxicity (*ToxinPred*)
stx2B					
NDTFTVKVAGKEYWT	Yes	No	0.8377	non-allergen	Non-toxin
AVLFALVSVNAMAAD	No	Yes	0.7232	Non-allergen	Non-toxin
stx1A					
MLRFVTVTAEALRFR	Yes	Yes	0.8854	Non allergen	Non-toxin
QRGFRTTLDDLSGRS	No	Yes	1.4316	Non-allergen	Non-toxin
fliC					
AGKITTETTSAGSAT	Yes	No	1.1663	Non allergen	Non-toxin
GTAVNVNSAGKITTE	No	Yes	1.1162	Non allergen	Non-toxin
escC					
EISVSIIDVDANDLQ	Yes	Yes	1.2454	Non allergen	Non-toxin
QIEISVSIIDVDAND	Yes	Yes	1.2617	Non-allergen	Non-toxin
eae					
GGSNIYKVTARAYDR	No	Yes	0.6885	Non allergen	Non-toxin
SVVIKATSGDKQTVS	Yes	No	1.3934	Non-allergen	Non-toxin

3.3 | Continuous and discontinuous B cell epitopes

Bepipred 2.0, IEDB and SVMTrip web servers were used to identify the continuous B-cell epitopes (Appendix S3). All predicted epitopes from the three servers were aligned with the Unipro UGENE Mafft algorithm (Figure 2) to get the consensus epitope used for designing the diagnostic kit. The discontinuous B cell epitopes were predicted with the ElliPro server, and six of them were selected whose scores were between 0.59 and 0.89 (Table S2).

3.4 | HTL prediction and screening of IL4 and IL10-inducing epitopes

The predicted HTL epitopes with strong binding affinity were selected, while those with low binding affinity were excluded (Table 2 and Table S3). Seventeen strong binders were IL4 inducers, and 11 were IL10 inducers HTL.

3.5 | Chimeric construct of the detection kit

The best epitope candidates were used for the construction of the detection kit; the B cell and HTL epitopes were joined with linker sequences (Figure 3).
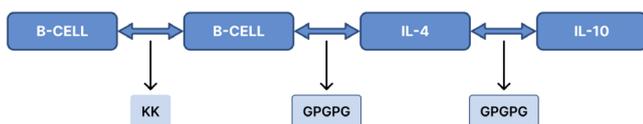

FIGURE 3 Chimeric construct of the peptide based kit.

3.6 | Antigenicity, solubility and physicochemical properties of the chimeric construct

The result from Vaxijen v2.0 server shows that the primary chimeric build is likely to be effective, with an antigenic score of 1.1123, above the 0.5 minimum threshold. Protein-Sol predicted the chimeric construct as soluble (0.623) by using the information on the solubility of *E. coli* proteins in cell-free expression systems. The protein construct comprises 295 amino acids in length, a molecular weight of 29.876 kDa, and a 9.06 theoretical protrusion index (pI) for determining solubility in an electric field, according to output from the Expasy Protparam web server. The protein is stable, as shown by the 14.82 instability index. With a grand average of hydropathicity (GRAVY) of −0.261 and an aliphatic index of 75.05, this protein construct demonstrates thermostability across a broad temperature range. The GRAVY values measure protein hydrophilicity, and a higher positive score denotes a more hydrophobic protein. As a result, the primary construct's low score indicates that it has a strong hydrophilicity and a propensity to interact with nearby water molecules. There were 30 positively charged residues (Arg + Lys) and 26 negatively charged residues (Asp + Glu). Mammalian reticulocytes, in vitro, 4.4 h; yeast, in vivo, >20 h; and >10 h (*E. coli*, in vivo) were computed as the chimeric construct half-life.

3.7 | Codon optimization and in-silico cloning

The codon adaptation was carried out in line with the expression host using the Java Codon Adaptation Tool (JCat), which converted the protein sequence into a nucleotide sequence. The revised sequence has a CAI value and a GC% of 1.0 and 53.33, respectively. In the *E. coli* expression vector pET-28a (+), the chimeric construct's optimized sequence
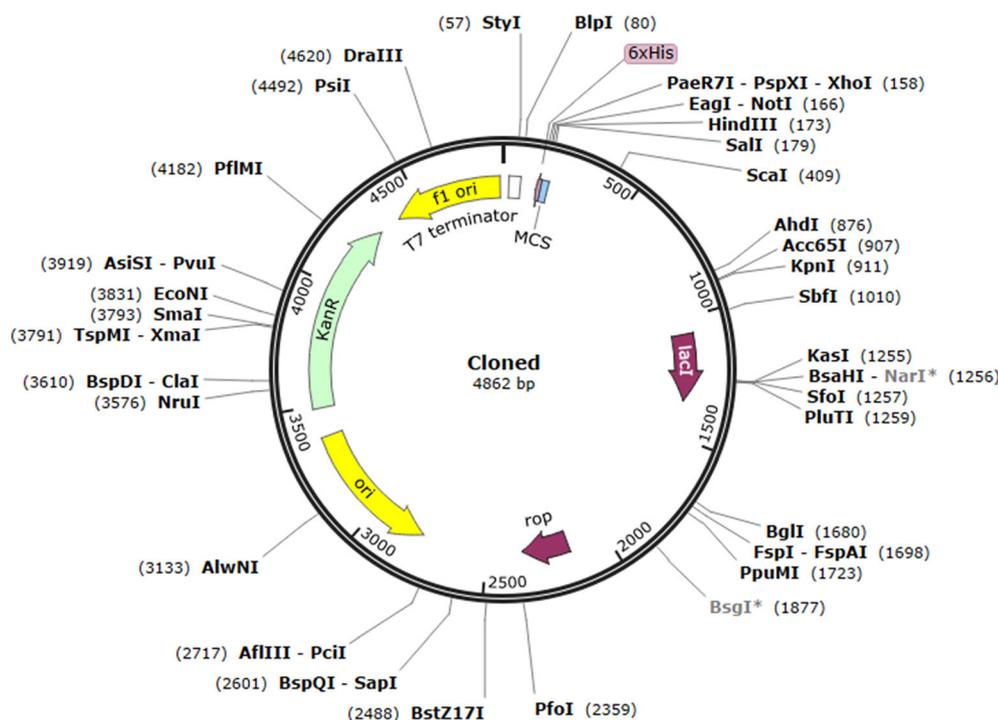

FIGURE 4 Optimized chimeric construct and pET28a(+).

(Improved DNA) was inserted between the Eco53kI and EcoRV restriction enzyme sites using. The clone's overall size was 4.862 kbp (Figure 4).

3.8 | Secondary and tertiary structure

The result from SOPMA shows that the construct has α-helices: 16.61%, β-turns: 5.42%, extended strands: 33.22%, and random coils: 44.75%. The significant proportion of extended strands and random coils of the chimeric construct raises the possibility that the protein might develop antigenic epitopes. Also, the 3D structural model of the chimeric construct was predicted by Alphafold 2.0 server. The models were ranked according to their
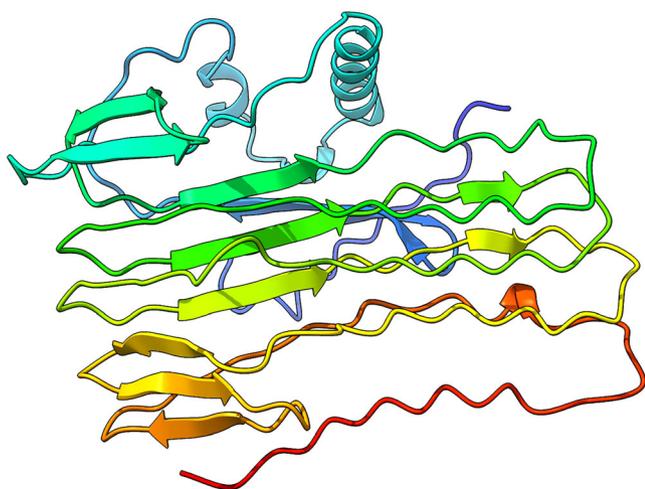
various confidence level scores. This study used the model with the highest confidence level for further analysis.

3.9 | Tertiary structure validation and refinement

The GalaxyRefine server received the tertiary structure model with the highest confidence score. After the submittal, five improved models were provided for the intended build. The model (Figure 5) with the best-fit parameters was chosen for further research. These parameters included GDTHA of 0.8847, RMSD of 0.600, MolProbity of 1.250, clash score of 4.7, Poor rotamers of 0.4, and Rama preferred of 98.0. After refining, ProSA-web and Procheck servers were used to verify the model's quality. The refined tertiary structure's energies as a function of amino acid sequence position illustrate the local model quality, and the overall model quality gave a Z-score of −4.39 (Figure 6). Meanwhile, Procheck's Ramachandran plot analysis reveals that 93.8%, 5.8%, and 0.4% of all residues were in preferred, additional allowed, and disallowed regions, respectively (Figure 7).

3.10 | Molecular docking with antigenic genes

HDOCK server yielded ten complex models with different docking scores for each antigenic gene and construct. The complex model with a more negative docking score and the highest confidence scores that cut across each set of docking results was selected. The docking and confidence score for each antigenic gene: stx2B (−239.13, 0.8560), stx1A (−301.68, 0.9451), fliC (−259.55, 0.8994), escC (−235.89, 0.8478), and eae (−293.88, 0.9467). The complex was then viewed with UCSF ChimeraX (Figure 8 and Appendix S5).
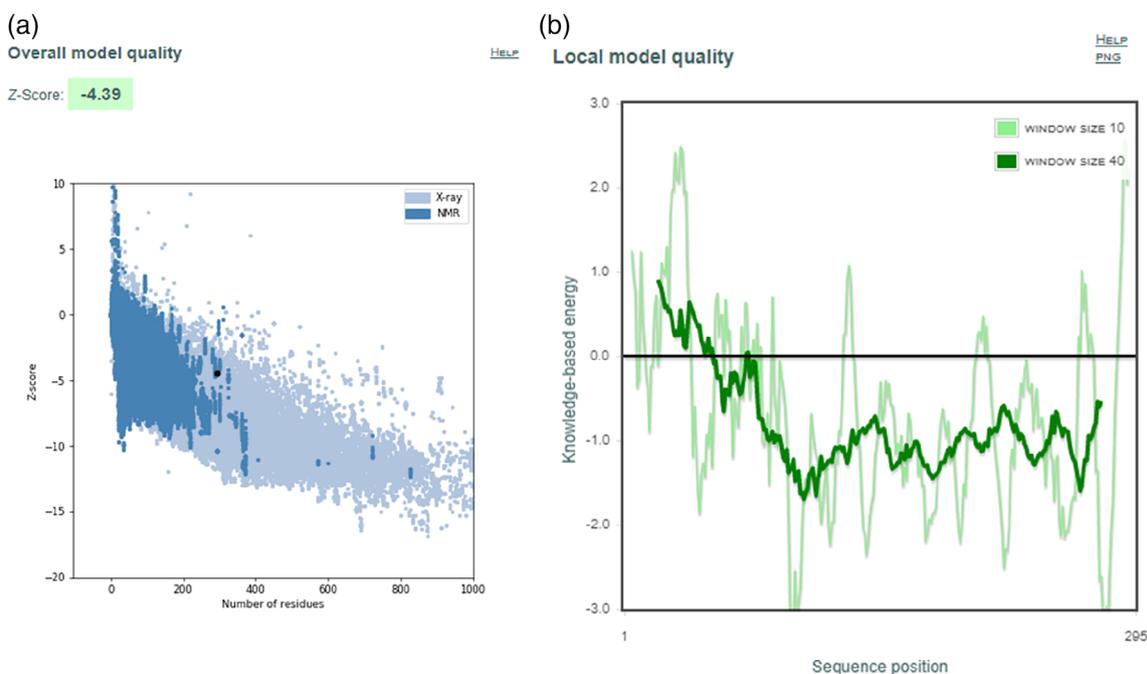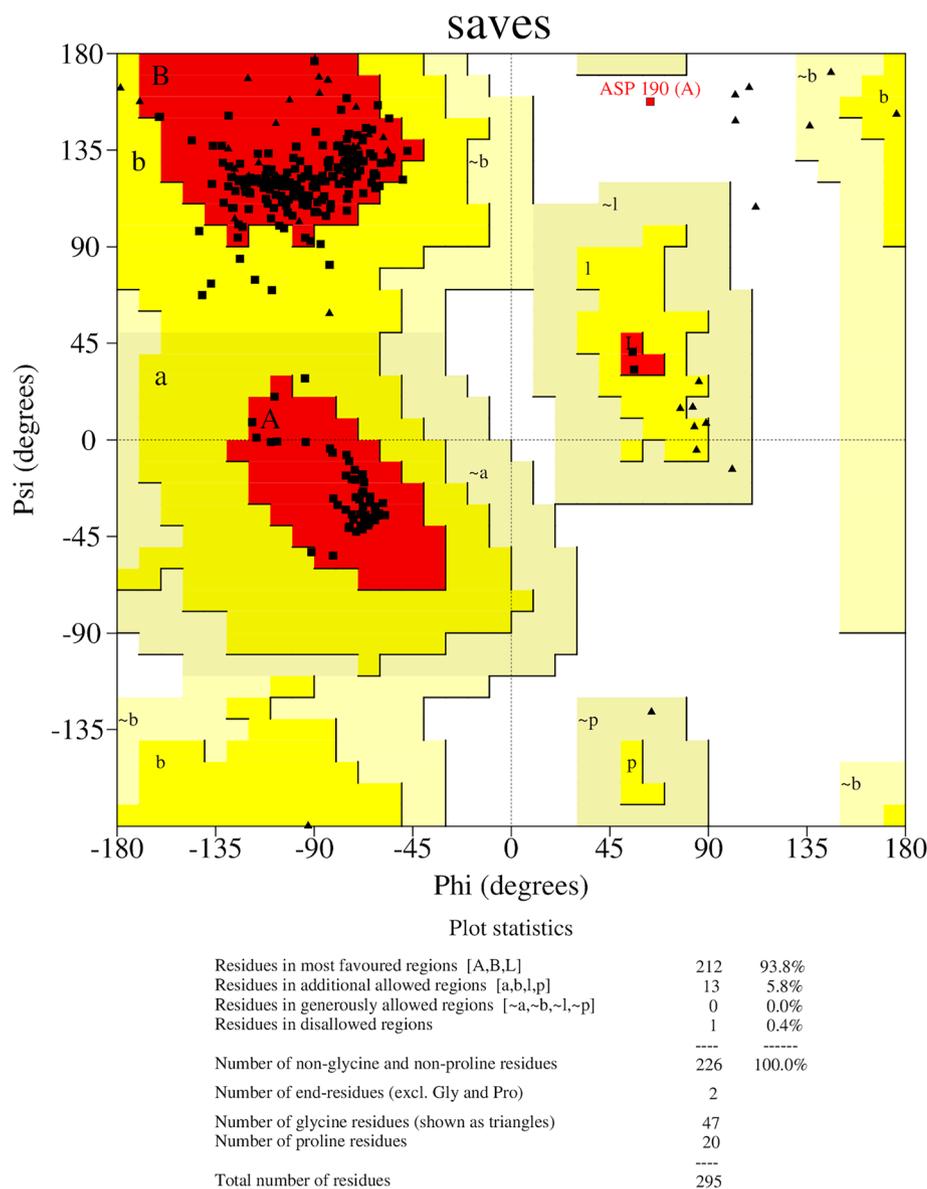

FIGURE 5 Refined tertiary protein structure, GalaxyRefine.

FIGURE 6 ProSA-web refined model validation result: (a) overall model quality with Z-score and (b) local model quality.

Ramachandran Plot
saves

FIGURE 7 Ramachandran plot of the refined tertiary structure.

Plot statistics

Residues in most favoured regions [A,B,L]	212	93.8%
Residues in additional allowed regions [a,b,l,p]	13	5.8%
Residues in generously allowed regions [~a,~b,~l,~p]	0	0.0%
Residues in disallowed regions	1	0.4%
	----	------
Number of non-glycine and non-proline residues	226	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	47	
Number of proline residues	20	

Total number of residues	295	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms
and R-factor no greater than 20%, a good quality model would be expected
to have over 90% in the most favoured regions.

4 | DISCUSSION

Foodborne illnesses are a prevalent concern in both developed and developing nations. The inaugural global burden study by the World Health Organization (WHO) reported that roughly 600 million individuals fall ill, and 33 million deaths result annually due to unsafe food and water (WHO, 2015). The growing consumer trend for pure, clean, safe food represents a substantial and continuing demand. Consumers' increasing concerns and evolving needs serve as a primary driver, creating pressure on decision-makers, food producers, and all individuals engaged in the food manufacturing process, including diagnostic and detection technology providers, to enhance food safety standards (Wiklund et al., 2002). Early detection represents an optimum preventive strategy to minimize the infection. That is why organizations like the WHO provided scientific assessments to control *E. coli* O157:H7, a significant pathogenic *E. coli* strain in food. Many diagnostic approaches have been used over time, such as emerging through-put methods that have offered numerous potential advantages, including isothermal DNA amplification, surface-enhanced Raman spectroscopy (SERS), biosensors, rapid paper-based diagnosis and digital smartphone-based methods. Additionally, commercially available testing kits are designed to detect *E. coli* O157:H7. Although few of these methods appear to meet the ASSURED criteria recommended by the WHO for an ideal detection test kit, further research
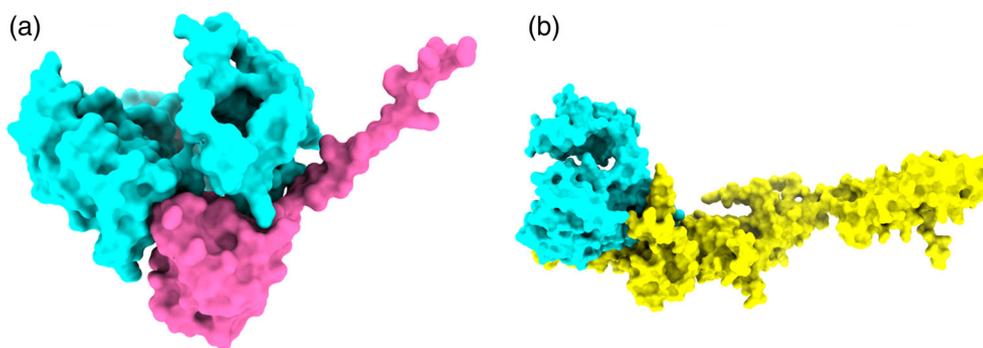
FIGURE 8 Diagrammatic representation of the detection kit structure docking results with some antigenic target genes. (a) The binding interaction between the diagnostic kit structure (cyan) and the Stx2B gene (pink). (b) The binding interaction between the diagnostic kit structure (cyan) and the escC gene (yellow). Other docked peptide result images are shown in Figure S2.

is needed to develop new techniques that can create a simple, compact, and portable device to facilitate early and rapid outbreak surveillance (Rani et al., 2021). With this in mind, the design of a diagnostic kit for detecting *E. coli* O157:H7 in food was proposed, specifically by recruiting immunoinformatic tools.

Studies have proven that the cell adhesion apparatus and the toxins secreted by this organism are the major cause of its pathogenicity (Najafabadi et al., 2021). The *E. coli* O157:H7 antigenic genes and encoding region sequences are *stx1*: Shiga toxin 1, *stx2*: Shiga toxin 2, *escC*: T3SS-type III secretion system, *eae*: Intimin, *fliC*: Flagellar antigen H7. They were screened to choose potential candidates for the diagnostic kit construct. The selected antigenic proteins must have epitopes in their exo-membrane sequence length for exposure and easy detection for antibody–antigen interaction (Ebrahimi et al., 2020). Continuous and discontinuous B-cell epitopes were predicted with Bepipred 2.0, IEDB, SVMTrip and Ellipro. Unipro UGENE was used to obtain a consensus for the epitopes of a particular gene. The selected peptides had 12–20 amino acids residue to get a reasonable molecular weight for the diagnostic kit design. After that, they were evaluated for antigenicity and stability to ensure they could stimulate immune response. Strong binding epitopes from HTLs that exhibit the ability to induce interleukin 4 (IL-4) and interleukin 10 (IL-10) were identified and selected (Oluwagbemi et al., 2022). In contrast, non-inducing epitopes were excluded from the study (Yu et al., 2017).

The criteria for selecting linkers included ensuring that the chosen linkers would increase B–cell stability and GPGPG would increase construct folding and biological activities (Chen et al., 2013). The development of linker peptides for synthetic protein fusion was inspired by research on naturally occurring multiple-domain proteins (Chen et al., 2013). According to research, the linkers utilized in this design are crucial for establishing an extended conformation (flexibility), separating functional domains, and folding proteins, all of which contribute to a more stable protein structure (Nezafat et al., 2014). The final chimeric protein construct consists of 295 amino acids with a molecular weight of 29.876 kDa, with other physicochemical features that suggest that the construct is stable (instability index of 14.82 and aliphatic index of 75.05), soluble (0.623) and hydrophilic

(GRAVY value: −0.261) with an estimated half-life of 4.4 h (mammalian reticulocytes, in vitro), >20 h (yeast, in vivo) and >10 h (*E. coli*, in vivo). The construct was also predicted to be very antigenic, with an antigenicity score of 1.1. The chimeric protein sequence was optimized, resulting in a GC content of 53.33% with a codon adaptation index (CAI) of 1.0. This is a favorable indication of the construct's stability and expression in E. coli. In-silico cloning was carried out using *E. coli* strain K-12 as the expression vector since it is widely adopted for increasing the expression and output of recombinant proteins (gene of interest) (Rosano & Ceccarelli, 2014). Therefore, the adapted codon sequence of the chimeric construct (fragment) was cloned in the *E. coli* K-12 pET28a(+) vector at Eco53kI and EcoRV restriction sites. Secondary structure analysis of the protein from the SOPMA server calculates that the composition of alpha (α) helices, extended strand, random coils, and beta (β) turn as 16.61%, 33.22%, 44.75%, and 5.42%, respectively. The diagnostic kit's tertiary structure was predicted using the Alphafold 2.0 server and refined with the GalaxyRefine server as the most relevant algorithms were employed. The selected model shows the best-fit parameters. The Ramachandran plot and ProSA-web statistics prove the quality of the tertiary construct; 93.8% of the residues were in the favored region, with only 0.4% in the disallowed region, as this confirms an ideal result and a Z-score of −4.39, which is the range of the identified structure topologically respectively.

The strong binding affinity of the diagnostic kit complex with receptors (3D-structure of the antigenic genes) was predicted by the HDOCK server. From the docking result, it was evident that all the antigenic genes had good docking scores; *stx1A* protein had the highest docking score of −301.68, then *eae* (Intimin) with a sore of −293.88, *fliC* with −259.55 score, *stx2B* with −239.13, and *escC* with the least docking score of −235.89 (Figure S2). *Stx* gene encodes for Shiga-toxins, a protein synthesis inhibitor, have been extremely sensitive in detecting E. coli O157:H7 (Rani et al., 2021), which informs its high docking score. The *Eae* gene encodes for Intimin, which serves as a ligand for epithelial cell adhesion. It has also been shown to stimulate mucosal Th1 immune responses and intestinal crypt hyperplasia

(Kaper et al., 2004). The observed affinity in the *fliC* gene could be attributed to its presence in other closely related bacterial strains (Beutin et al., 2005). Even though the *EscC* gene has been identified as one of the main virulence traits of EHEC (Braz et al., 2020), it was shown to have the least docking score compared to other genes but was reasonably higher than −200.

5 | CONCLUSION

The utilization of bioinformatic and computational methods in detecting *E. coli* O157:H7 holds great potential for the creation of new diagnostic tools for this water-borne pathogen. Based on the findings from epitope peptide analysis, *Z*-score, and Ramachandran plot analysis, it can be concluded that the final construct and clone have successfully met the design requirements for the development of a lateral flow kit, which has the potential to provide fast and efficient detection of *E. coli* O157:H7. This could ultimately reduce mortality rates by improving treatment options and expediting decision-making. However, it is essential to acknowledge that additional validation through in vitro and in vivo techniques is necessary to further confirm this detection approach's efficacy and accuracy.

CONFLICT OF INTEREST STATEMENT
The authors reported no conflict of interest.

DATA AVAILABILITY STATEMENT
All data analyzed in this study will be madeavailable upon request.

ORCID
Elijah Kolawole Oladipo https://orcid.org/0000-0002-9646-5122
Jerry Ayobami Oluwasegun https://orcid.org/0000-0003-1454-2663
Tolu David Oladunni https://orcid.org/0000-0002-7437-0763
Victoria Oluwakemi Owoeye https://orcid.org/0009-0003-1349-9007
Olukayode Ifeanyi Obisanya https://orcid.org/0000-0002-8025-8110
Oluwatosin Tumininu Olasinde https://orcid.org/0009-0000-4680-9428
Helen Onyeaka https://orcid.org/0000-0003-3846-847X

REFERENCES
Batchelor, M., Prasannan, S., Daniell, S., Reece, S., Connerton, I., Bloomberg, G., Dougan, G., Frankel, G., & Matthews, S. (2000). Structural basis for recognition of the translocated intimin receptor (Tir) by intimin from enteropathogenic Escherichia coli. *The EMBO Journal*, 19(11), 2452–2464.

Beutin, L., Strauch, E., Zimmermann, S., Kaulfuss, S., Schaudinn, C., Männel, A., & Gelderblom, H. R. (2005). Genetical and functional investigation of fliC genes encoding flagellar serotype H4 in wildtype strains of Escherichia coli and in a laboratory E. coli K-12 strain expressing flagellar antigen type H48. *BMC Microbiology*, 5(1), 1–11.

Braz, V. S., Melchior, K., & Moreira, C. G. (2020). Escherichia coli as a multifaceted pathogenic and versatile bacterium. *Frontiers in Cellular and Infection Microbiology*, 10, 548492.

Brooks, J. T., Sowers, E. G., Wells, J. G., Greene, K. D., Griffin, P. M., Hoekstra, R. M., & Strockbine, N. A. (2005). Non-O157 Shiga toxin-producing Escherichia coli infections in the United States, 1983–2002. *The Journal of Infectious Diseases*, 192(8), 1422–1429.

Bunger, J. C., Melton-Celsa, A. R., Maynard, E. L., & O'Brien, A. D. (2015). Reduced toxicity of Shiga toxin (Stx) type 2c in mice compared to Stx2d is associated with instability of Stx2c holotoxin. *Toxins*, 7(6), 2306–2320.

Chen, X., Zaro, J., & Shen, W. C. (2013). Fusion protein linkers: Effects on production, bioactivity, and pharmacokinetics. In *Fusion protein technologies for biopharmaceuticals: Applications and challenges* (pp. 57–73). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118354599.ch4

Clifford, J. N., Høie, M. H., Deleuran, S., Peters, B., Nielsen, M., & Marcatili, P. (2022). BepiPred-3.0: Improved B-cell epitope prediction using protein language models. *Protein Science*, 31(12), e4497.

Deisingh, A. K., & Thompson, M. (2004). Strategies for the detection of Escherichia coli O157:H7 in foods. *Journal of Applied Microbiology*, 96(3), 419–429.

Dhanda, S. K., Gupta, S., Vir, P., & Raghava, G. P. S. (2013). Prediction of IL4 inducing peptides. *Clinical and Developmental Immunology.*, 2013, 1–9.

Dimitrov, I., Bangov, I., Flower, D. R., & Doytchinova, I. (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.

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.

Fatima, R., & Aziz, A. (2023). Enterohemorrhagic Escherichia Coli. [Updated 2023 Feb 22]. In: *StatPearls* [Internet]. Treasure Island (FL): StatPearls Publishing. Available from: https://www.ncbi.nlm.nih.gov/books/NBK519509/

García-Angulo, V. A., Kalita, A., Kalita, M., Lozano, L., & Torres, A. G. (2014). Comparative genomics and immunoinformatics approach for the identification of vaccine candidates for enterohemorrhagic Escherichia coli O157:H7. *Infection and Immunity*, 82(5), 2016–2026.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., & Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, 31(13), 3784–3788.

Gehring, A. G., & Tu, S. I. (2005). Enzyme-linked immunomagnetic electrochemical detection of live Escherichia coli O157:H7 in apple juice. *Journal of Food Protection*, 68(1), 146–149.

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.

Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., & Raghava, G. P. S. (2015). Peptide toxicity prediction. In *Methods in molecular*

biology (Vol. 1268, pp. 143–157). Springer Science+Business Media. https://doi.org/10.1007/978-1-4939-2285-7_7

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 (Oxford, England)*, *33*(19), 3098–3100. https://doi.org/10.1093/bioinformatics/btx345

Hennebert, E., Maldonado, B., Ladurner, P., Flammang, P., & Santos, R. (2015). Experimental strategies for the identification and characterization of adhesive proteins in animals: A review. *Interface Focus*, *5*(1), 20140064.

Heo, L., Park, H., & Seok, C. (2013). GalaxyRefine: Protein structure refinement driven by side-chain repacking. *Nucleic Acids Research*, *41*, W384–W388. https://doi.org/10.1093/nar/gkt458

Hallgren, J., Tsirigos, K. D., Pedersen, M. D., Almagro Armenteros, J. J., Marcatili, P., Nielsen, H., Krogh, A., & Winther, O. (2022). DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. *BioRxiv*, 2022-04. https://www.biorxiv.org/content/10.1101/2022.04.08.487609v1.full

Kalita, P., Lyngdoh, D. L., Padhi, A. K., Shukla, H., & Tripathi, T. (2019). Development of multi-epitope driven subunit vaccine against Fasciola gigantica using immunoinformatics approach. *International Journal of Biological Macromolecules*, *138*, 224–233.

Kaper, J., Nataro, J., & Mobley, H. (2004). Pathogenic Escherichia coli. *Nature Reviews Microbiology*, *2*(2), 123–140.

Karmali, M. A., Gannon, V., & Sargeant, J. M. (2010). Verocytotoxin-producing Escherichia coli (VTEC). *Veterinary Microbiology*, *140*(3–4), 360–370. https://doi.org/10.1016/j.vetmic.2009.04.011

Laskowski, R. A., & Thornton, J. M. (2021). PDBsum extras: SARS-CoV-2 and AlphaFold models. *Protein Science*, *31*, 283–289.

Loge, F. J., Thompson, D. E., & Call, D. R. (2002). PCR detection of specific pathogens in water: A risk-based analysis. *Environmental Science & Technology*, *36*(12), 2754–2759.

Mahajan, A., Currie, C. G., Mackie, S., Tree, J., McAteer, S., McKendrick, I., McNeilly, T. N., Roe, A., La Ragione, R. M., Woodward, M. J., Gally, D. L., & Smith, D. G. (2009). An investigation of the expression and adhesin function of H7 flagella in the interaction of Escherichia coli O157:H7 with bovine intestinal epithelium. *Cellular Microbiology*, *11*(1), 121–137. https://doi.org/10.1111/j.1462-5822.2008.01244

Marks, H. M., Tohamy, S. M., & Tsui, F. (2013). Modeling uncertainty of estimated illnesses attributed to non-O157:H7 Shiga toxin-producing Escherichia coli and its impact on illness cost. *Journal of Food Protection*, *76*(6), 945–952. https://doi.org/10.4315/0362-028X.JFP-12-409

Mellmann, A., Bielaszewska, M., Köck, R., Friedrich, A. W., Fruth, A., Middendorf, B., Harmsen, D., Schmidt, M. A., & Karch, H. (2008). Analysis of collection of hemolytic uremic syndrome-associated enterohemorrhagic *Escherichia coli*. *Emerging Infectious Diseases*, *14*(8), 1287–1290.

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.

Nagpal, G., Usmani, S. S., Dhanda, S. K., Kaur, H., Singh, S., Sharma, M., & Raghava, G. P. (2017). Computer-aided designing of immunosuppressive peptides based on IL-10 inducing potential. *Scientific Reports*, *7*(1), 42851.

Najafabadi, Z. S., Nazarian, S., Kargar, M., & Kafilzadeh, F. (2021). Designing of a chimeric protein contains StxB, intimin and EscC against toxicity and adherence of enterohemorrhagic Escherichia coli O157:H7 and evaluation of serum antibody titers against it. *Molecular Immunology*, *134*, 218–227.

Nezafat, N., Ghasemi, Y., Javadi, G., Khoshnoud, M. J., & Omidinia, E. (2014). A novel multi-epitope peptide vaccine against cancer: an in silico approach. *Journal of Theoretical Biology*, *349*, 121–134.

Nguyen, Y., & Sperandio, V. (2012). Enterohemorrhagic Escherichia coli (EHEC) pathogenesis. *Frontiers Cell Infectious Microbiology*, *2*, 90. https://doi.org/10.3389/fcimb.2012.00090

Oladipo, E. K., Ajayi, A. F., Odeyemi, A. N., Akindiya, O. E., Adebayo, E. T., Oguntomi, A. S., Oyewole, M. P., Jimah, E. M., Oladipo, A. A., Ariyo, O. E., Oladipo, B. B., & Oloke, J. K. (2020). Laboratory diagnosis of COVID-19 in Africa: Availability, challenges and implications. *Drug Discoveries & Therapeutics*, *14*(4), 153–160. https://doi.org/10.5582/ddt.2020.03067

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 & dynamics*, *10*(7), 1–8. https://doi.org/10.1080/07391102.2022.2111358

Oluwagbemi, O., Oladipo, E., Kolawole, O., Oloke, J., Adelusi, T., Irewolede, B., Dairo, E., Ayeni, A., Kolapo, K., Akindiya, O., Oluwasegun, J., Oluwadara, B., & Fatumo, S. (2022). Bioinformatics, computational informatics, and modeling approaches to the design of mRNA COVID-19 vaccine candidates. *Computation*, *10*(7), 117. https://doi.org/10.3390/computation10070117

Ponomarenko, J., Bui, H. H., Li, W., Fusseder, 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), 1–8.

Rani, A., Ravindran, V. B., Surapaneni, A., Mantri, N., & Ball, A. S. (2021). Trends in point-of-care diagnosis for Escherichia coli O157:H7 in food and water. *International Journal of Food Microbiology*, *349*, 109233. https://doi.org/10.1038/s41598-021-81312-6

Reynisson, B., Barra, C., Kaabinejadian, S., Hildebrand, W. H., Peters, B., & Nielsen, M. (2020). Improved prediction of MHC II antigen presentation through integration and motif deconvolution of mass spectrometry MHC eluted ligand data. *Journal of Proteome Research*, *19*(6), 2304–2315.

Ribot, E. M., Fair, M. A., Gautom, R., Cameron, D. N., Hunter, S. B., Swaminathan, B., & Barrett, T. J. (2006). Standardization of pulsed-field gel electrophoresis protocols for the subtyping of Escherichia coli O157:H7, Salmonella, and Shigella for PulseNet. *Foodborne Pathogens and Disease*, *3*(1), 59–67. https://doi.org/10.1089/fpd.2006.3.59

Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in Escherichia coli: Advances and challenges. *Frontiers in Microbiology*, *5*, 172.

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. https://doi.org/10.1186/s12934-019-1053-9

Weidmann, M., Faye, O., Faye, O., Abd El Wahed, A., Patel, P., Batejat, C., Manugerra, J. C., Adjami, A., Niedrig, M., Hufert, F. T., & Sall, A. A. (2018). Development of mobile laboratory for viral hemorrhagic fever detection in Africa. *The Journal of Infectious Diseases*, *218*(10), 1622–1630.

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.

Wiklund, P., Orava, M., Brännback, M., & De Heer, A. J. (2002). The present state of the field of food diagnostics. Innomarket: Turku School of Economics and Business Administration Department of Marketing. *Technical Report*, *10*, 10–13.

Wong, A. R., Pearson, J. S., & Bright, M. D. (2011). Enteropathogenic and enterohaemorrhagic Escherichia coli: Even more subversive elements. *Molecular Microbiology*, *80*(6), 1420–1438.

World Health Organization: WHO. (2015). *Food safety*. https://www.who.int/news-room/fact-sheets/detail/food-safety. Accessed February 10, 2023.

Yan, Y., Tao, H., He, J., & Huang, S. Y. (2020). The HDOCK server for integrated protein–protein docking. *Nature Protocols*, *15*(5), 1829–1852. https://doi.org/10.1038/s41596-020-0312-x

Yao, B., Zheng, D., Liang, S., & Zhang, C. (2020). SVMTriP: A method to predict B-cell linear antigenic epitopes. In *Immunoinformatics* (pp. 299–307). Humana.

Yu, Y., Chen, Y., Wang, F. L., Sun, J., Li, H. J., & Liu, J. M. (2017). Cytokines interleukin 4 (IL-4) and interleukin 10 (IL-10) gene polymorphisms as potential host susceptibility factors in virus-induced encephalitis. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research*, *23*, 4541–4548.

Zhao, S., Wang, S., Zhang, S., Liu, J., & Dong, Y. (2018). State of the art: Lateral flow assay (LFA) biosensor for on-site rapid detection. *Chinese Chemical Letters*, *29*(11), 1567–1577.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.