Iraqi Journal of Science, 2026, Vol. xx, No. x, pp: xx
DOI: 10.24996/ijs.2026.67.2.16

ISSN: 0067-2904

Structural modelling and *in silico* analysis of a thiol-activated cytolysin produced by *Bacillus thuringiensis*

Husam Sabah Auhim [1]*, Nisreen Hadi Odaa[1], Ismail Taban[2]
[1] *Department of Biology, College of Science, University of Baghdad, Baghdad, Iraq.*
[2] *Misurata, Biotechnology Research Center, Libya*

Received: 4/ 11/2024 Accepted: 2/ 2/2025 Published: xx

Abstract

Bacillus thuringiensis (Bt) is a soil-dwelling, Gram-positive bacterium that produces proteins with specific toxic effects against invertebrate organisms. These toxins specifically target insects. This study examined a hypothetical Bt protein for its potential toxicity to mammals. The hypothetical protein was analyzed using various bioinformatics tools, including NCBI BLAST and conserved domain tools. The physicochemical properties were assessed using ExPasy ProtParam, and homology models were generated through automated services. The analysis revealed that the hypothetical protein is a thiol-activated cytolysin, a pore-forming toxin that targets mammalian cells with a signature motif. The toxin has a molecular weight of 52.6 kDa, and its instability index is 25.73, indicating it is stable under physiological conditions. Furthermore, the toxin is expected to exhibit thermal stability, as evidenced by its aliphatic index of 79.52. Three homology modeling services were used, homology modeling identified anthrolysin O, a thiol-activated cytolysin with 97.53% sequence identity, as the closest template. The Swiss model demonstrated superior validation compared to others. The protein folds into four functional domains characteristic of thiol-activated cytolysins: domain 3 likely forms beta-barrel channels, domains 1 and 2 are involved in oligomerization, and domain 4 recognizes cholesterol. Docking analysis revealed hydrophobic binding interactions with 8 of 11 residues in the cholesterol signature motif, including a hydrogen bond with Gly474 (2.29 Å). In conclusion, Bt produces a thermostable cholesterol-dependent cytolysin that shows strong similarity to anthrolysin O and perfringolysin O.

Keywords: *Bacillus thuringiensis*, Cholesterol dependent cytolysin, Protein modelling, Pore-forming toxin, Thiol-activated cytolysin.

النمذجة الهيكلية والتحليل السيلكي للسايتوليسين المنشط بالثايول المنتج من بكتريا *Bacillus thuringiensis*

حسام صباح أوهيم*[1], نسرين هادي عودة[1] , أسماعيل طابان [2]
[1] قسم علوم الحياة , كلية العلوم , جامعة بغداد, بغداد , العراق
[2] مركز بحوث التقنيات الاحيائية, مصراتة , ليبي

الخلاصة

* Email: husam.auhim@sc.uobaghdad.edu.iq

بكتيريا *Bacillus thuringiensis*، والمعروفة باسم (Bt) هي بكتيريا إيجابية لصبغة الجرام تعيش في التربة وتنتج بروتينات ذات تأثيرات سامة محددة ضد الكائنات اللافقارية. تستهدف هذه السموم الحشرات على وجه التحديد. فحصت هذه الدراسة بروتين *Bacillus thuringiensis* افتراضيًا لمعرفة سميته المحتملة للثدييات. تم تحليل البروتين الافتراضي باستخدام أدوات المعلوماتية الحيوية المختلفة، بما في ذلك NCBI BLAST وأداة NCBI conserved domains. تم تقييم الخصائص الفيزيائية والكيميائية باستخدام ExPasy ProtParam، وتم إنشاء نماذج التماثل من خدمات نمذجة بنية البروتين الآلية. كشف التحليل أن البروتين الافتراضي هو سيتوليزين منشط بالثيول، وهو سم مكون للمسام يستهدف خلايا الثدييات بنمط مميز. يبلغ الوزن الجزيئي للسم 52.6 كيلو دالتون، ومؤشر عدم استقراره 25.73، مما يشير إلى أنه مستقر في ظل الظروف الفسيولوجية. علاوة على ذلك، من المتوقع أن يُظهر السم استقرارًا حراريًا، كما يتضح من مؤشره الأليفاتي البالغ 79.52. تم استخدام ثلاث خدمات نمذجة متماثلة ، وحددت نمذجة التجانس الأنثروليسين O، وهو سيتوليزين منشط بالثيول بنسبة هوية تسلسل 97.53٪، باعتباره القالب الأقرب. أظهر النموذج السويسري دقة متفوقة مقارنة بالآخرين. ينطوي البروتين على أربعة مجالات وظيفية مميزة للسيتوليزينات المنشطة بالثيول: المجال 3 من المحتمل أن يشكل قنوات بيتا، والمجالان 1 و 2 يشاركان في تكوين الأوليغومر، والمجال 4 يتعرف على الكوليسترول. كشف تحليل الالتحام عن تفاعلات ارتباط كارهة للماء مع 8 من 11 حامض اميني في signature motif المميز للكوليسترول ، بما في ذلك رابطة هيدروجينية مع Gly474 (2.29 Å). في الختام، ينتج Bt سيتوليزين ثابتًا حراريًا معتمدًا على الكوليسترول ويظهر تشابهًا قويًا مع الأنثروليسين O والبيرفرينجولايسين O..

1. Introduction

The thiol-activated cytolysin family, also referred to as cholesterol-dependent cytolysin (CDCs), comprises a group of pore-forming toxins produced by several Gram-positive bacteria. These toxins are characterized by their ability to lyse cholesterol-containing membranes, reversible deactivation by oxidation, and binding to cholesterol. All of these toxins have a single cysteine residue (thiol group) in their C-terminal domain, which was established to be necessary for cholesterol binding [1]. Additionally, this family has an undecapeptide sequence (ECTGLAWEWWR) known as a signature motif; all the members belonging to this family share the same motif. This motif has an essential role in pore formation during toxicity [2]. These pore-forming toxins are initially produced as monomers (water-soluble 50-70 kDa), forming a complex of circular homo-oligomeric consisting of as many monomers (40 or more) when attached to the target cell. Following several conformational changes in the protein domains, a structure of β-barrel transmembrane (~250 Å in diameter, depending on the toxin) is formed and embedded into the target cell membrane. Once the toxin forms the pore (~250 Å in diameter), which is spacious enough to allow the loss of essential molecules for the cell survival, such as amino acids, nucleotides, small and large proteins, as well as ions Ca^{2+}, and Na^{+}, $K^{+}3$, the cell's integrity is compromised.

Bacillus thuringiensis known as (Bt), can synthesize various classes of toxins that target invertebrates, mainly insects, including a parasporal crystal containing crystal (Cry) and cytolytic (Cyt) toxins (also known as collectively δ-endotoxins) during sporulation [4]. Vegetative insecticidal proteins (Vip), and secreted insecticidal protein (Sip). Since they are pathogenic to insects, they have been used as biological pesticides [5]. Additionally, these toxins have been widely used in transgenic plants. Bt can produce various proteins, including 41.9-kDa toxin and beta exotoxins. Proteins related to cholesterol-dependent cytolysins have been reported in some *B. thuringiensis* strains, such as *B. thuringiensis* subsp. *Kurstaki* [6]. Interestingly, certain Bt toxins that lack insecticidal activity have been demonstrated to have anticancer activity. It produces proteins known as parasporins (PSs), which have cytotoxic effects on different human cancer cells. These proteins are categorized into six groups, PS1 to PS6, each with specific anticancer mechanisms. PS3 and PS6 are pore-forming toxins that

disrupt the plasma membrane of cancer cells. PS2, a member of β pore-forming toxins (aerolysin-type), is a promising therapeutic agent for cancer treatment owing to its potent and specific cytotoxicity towards human cancer cells without disrupting ordinary cells [7].
Understanding protein function, including their interactions with other molecules, necessitates knowledge of a protein's structure. X-ray crystallography techniques can be used to determine the molecular structure of a protein. However, in some cases, it is hard to determine the protein structure experimentally. Homology modelling can be used to predict a protein's molecular structure using a known experimental structure of a homologous protein as a template [8]. Molecular modelling is based on the fact that naturally occurring proteins with sequence identity usually have similar molecular structures [9]. Today, numerous bioinformatics tools have been developed to predict three-dimensional protein structures, and in many cases, these predictions closely resemble the actual structure. Predicted structures can be very useful tools for examining protein structure-function relationships [8].

The main goal of this study was to investigate the physicochemical characteristics of a novel thiol-activated cytolysin variant synthesised by *B. thuringiensis*. This was accomplished by utilizing several bioinformatics tools and constructing a protein model to facilitate structural analysis.

2. Materials and methods

2.1 Protein analysis

The whole genome sequence and isolate of *B. thuringiensis* were graciously provided by Professor Dr. Colin Berry from the School of Biosciences at Cardiff University in Cardiff, United Kingdom. The DNA sequence corresponding to a hypothetical protein was extracted from the complete genome and subsequently subjected to a comprehensive analysis of the protein sequence, encompassing structural modelling, functional annotation, and protein analysis. Initially, the DNA sequence was analyzed using the Basic Local Alignment Search Tool (BLAST) available at https://blast.ncbi.nlm.nih.gov/Blast.cgi. The blastx tool [10] was employed to convert the DNA sequence into a corresponding protein sequence. The NCBI conserved domains tool and InterPro were employed to ascertain the protein family and significant domains (https://www.ncbi.nlm.nih.gov/Structure/cdd/ wrpsb.cgi). The SignalP 5.0 Server, available at http://www.cbs.dtu.dk/services/SignalP/, was employed for signal peptide prediction. The physical parameters of the protein, including the number of amino acids, molecular weight, theoretical isoelectric point (pI), instability index, extinction coefficient, and aliphatic index, were determined using the expert protein analysis system (ExPASy) available at www.expasy.org [11]. The process of constructing models for protein structures involved the utilisation of automated services specifically designed for homology-modeling. These services, namely Swiss-model [12], I-TASSER [13], and Phyre [14], were employed in the model-building process. A Swiss-model structure assessment tool was used for model validation and structure evaluation [12]. PyMOL software was used for visualisation and graphical representations of protein structure.

2.2 Docking study

The validated model that outperformed the others was selected for docking using the Auto Dock Tool version 1.5.7 [15]. The docking calculations were performed using the default values. No water molecules were considered. 30 output poses were obtained for cholesterol.

3. Results

3.1 DNA sequence analysis

The result of the nucleotide BLAST search of a hypothetical protein revealed that the target DNA sequence has 99.85% sequence identity with *Bacillus weihenstephanensis* strain WSBC

10204 (GenBank: CP009746.1). This DNA sequence of *Bacillus weihenstephanensis* belongs to the anthrolysin O family (cholesterol-dependent cytolysin), a pore-forming toxin family member. The target DNA sequence was translated into a protein using blastx for protein sequence analysis.

3.2 Protein classification

The analysis of conserved domains revealed the presence of two conserved domains within the protein. The initial conserved domain identified was the thiol-activated cytolysin, followed by the thiol-activated cytolysin beta-sandwich domain. The latter domain has an immunoglobulin-like shape and is located at the C-terminus of the thiol-activated cytolysin protein. Additionally, the conserved domains tool and InterPro have classed the protein as belonging to the thiol-activated cytolysin family protein. This protein is known for its ability to generate pores and lyse host membranes that contain cholesterol. Moreover, InterPro has identified the final domain as perfringolysin O domain. Based on these findings, it can be inferred that the target protein exhibits characteristics typical of a thiol-activated cytolysin, belonging to the pore-forming toxin family. In order to differentiate between the hypothetical protein produced by the *B. thuringiensis* strain and other members of thiol-activated cytolysins, it has designated as thurinolysin.

3.3 The signature motif of thiol-activated cytolysin

The undecapeptide sequence, which is 11 amino acids (ECTGLAWEWWR), is the most highly conserved region in the primary structure of the thiol-activated cytolysins and is unknown as a signature motif for these toxins. Changes in this motif's primary structure significantly impact the pore-forming mechanism of the thiol-activated cytolysins, which use cholesterol as their receptor on the target cell [3]. Thurinolysin protein shares the same signature motif, as shown in Figure 1. This result confirms that the thurinolysin protein is a member of thiol-activated cytolysins.

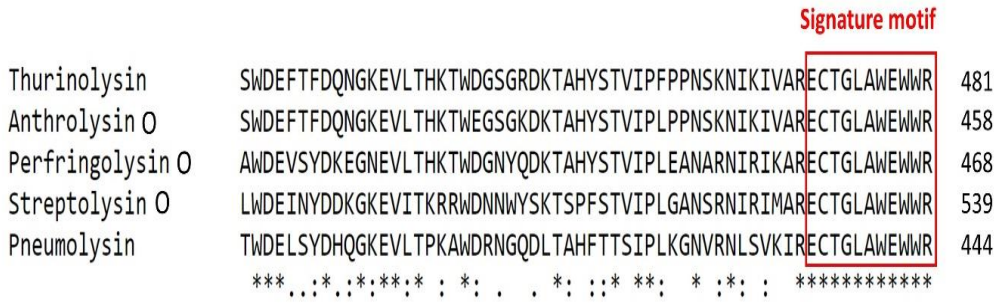

Figure 1: The undecapeptide sequence motif (signature motif, ECTGLAWEWWR, pointed by a red colour square) shared between the thiol-activated cytolysins. Anthrolysin O (Accession: WP_003163758.1), perfringolysin O (Accession: WP_003462918.1), streptolysin O (Accession: WP_010921831.1), pneumolysin (Accession: WP_088781280.1).

3.4 Physicochemical characteristics

The SignalP 5.0 server [16] was used to determine the signal peptide of thurinolysin protein. The findings of the study unveiled the specific location of the cleavage site occurring at residues 34 and 35. Furthermore, it was noted that the protein was secreted via the general secretion system known as (Sec-pathway). The first 34 residues were identified as the signal peptide. A total of 34 residues were eliminated from the sequence in order to generate the mature protein for subsequent investigation.

Different physicochemical properties of the mature thurinolysin protein were examined using the ExPASy ProtParam tool. The number of amino acids is 478, with a molecular weight of

52648.97 Da and a theoretical pI of 6.07. The extinction coefficient of thurinolysin protein is 62800 M^{-1} cm^{-1}, at 280 nm measured in water. The instability index was computed to be 25.73. The instability index of a protein lower than 40 is predicted as stable, while above that, it is unstable [17]. Based on this result, thurinolysin protein classifies as a stable protein. The AI of thurinolysin protein was 79.52, indicating that this protein is thermally stable as well as contains a high amount of hydrophobic amino acids.

3.5 Structural modelling of thurinolysin protein

3.5.1 Identification of a template

The first and most critical step in homology modeling is identifying the optimal template structure to serve as the basis for the model [18]. The simplest and most straightforward approach for this template identification process is to perform serial pairwise sequence alignments against sequence databases, such as those supported by the BLAST. Utilizing the BLAST tool from the ExPASy server to systematically search the PDB for similar structural templates is a well-established and commonly used approach in homology modeling. The selection of appropriate template structures is based on several factors, including length of the sequence alignment, E-value from the BLAST search (ideally E-value < 0.005 for a good match), percent sequence identity between the query and template, BLAST score. 26 different templates that have various % identities and E-values were identified. Thiol-activated cytolysin for *Bacillus anthracis* (Accession 3CQF_A) has the best E-value of zero, and the % identity was 97.53%. The phylogenetic tree result of crystal structures and the protein query (Unknown, yellow colour) placed the query protein in the same branch with *B. Anthracis,* as depicted in Figure 2.

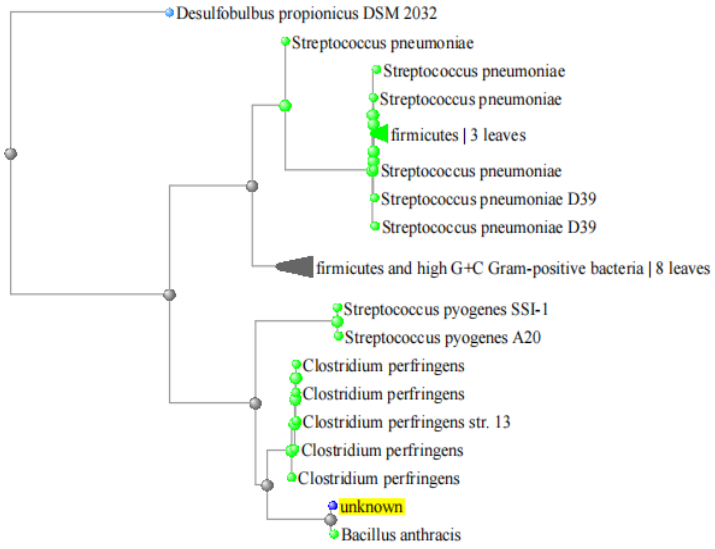

Figure. 2: Phylogenetic tree for templates and query protein (unknown)

3.5.2 Homology building

To obtain preliminary structural information on thurinolysin protein, an *in silico* modelling approach was employed. The thurinolysin protein was modeled using automated protein structure homology-modelling services. Three models were built by Swiss-model, I-TASSER and Phyre services, as depicted in Figure 3. First, the closest available homologues in the PDB as a template were searched using the above services. Results showed that all the templates belong to the thiol-activated cytolysin family protein produced by different bacteria. The template with the highest identity (97.53, BDP code: 3cqf) is a molecular structure of the anthrolysin O (thiol-activated cytolysin) produced by *B. anthracis*, which has been selected for model building.

The three models generally exhibit similar overall folding, with minor variations mainly observed in the N-terminal region, especially in the Phyre model. However, due to various algorithms used for protein structure prediction and building a model, distinct models may be generated for a given protein sequence [19]. Therefore, a high-quality model must be selected for further studies based on model validation.

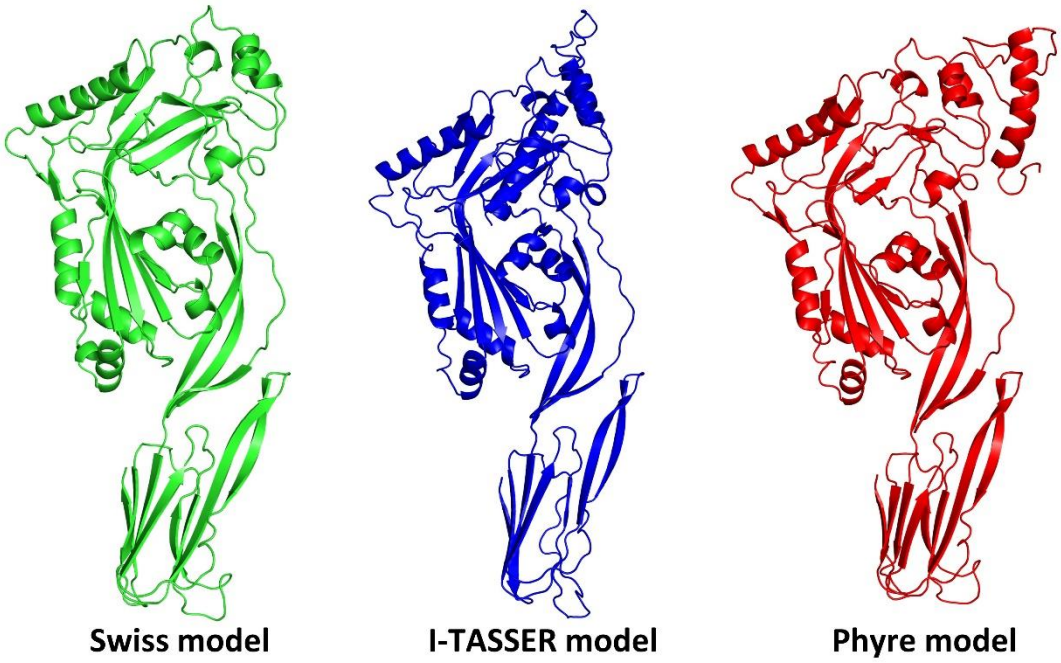

Figure 3: Thurinolysin models generated by different automated protein structure homology-modelling services.

3.5.3 Model validation

Numerous studies have analyzed the correlation between sequence identity and model accuracy, typically measured by metrics like root-mean-square deviation (RMSD) or global distance test scores. Generally, a trend of decreasing model accuracy is observed as sequence identity drops: At sequence identities above 50%, models tend to have RMSD values less than 2Å from the true structure. In the 30-50% sequence identity range, RMSD values typically fall between 2-4Å. Below 30% sequence identity, RMSD values can exceed 4Å, with larger variations in model quality [20]. The template used for building the thurinolysin model shares 97.53 sequence identity; therefore, the level of model accuracy should be high. However, models evaluation by the Swiss-model structure assessment tool shows that the Swiss model was the best among the others based on different validation criteria. In contrast, Phyre model has a lower quality with plenty of errors, as illustrated in Table 1. Consequently, the Swiss model was selected for structural analysis.

Table 1: Validation criteria for the evaluation of thurinolysin models.

Validation criteria	Swiss model	I-TASSER model	Phyre model	Ideal case
MolProbity Score	1.76	2.52	3.07	As low as possible
Clash Score	2.30	2.39	85.28	As low as possible
Ramachandran Favored	95.60%	79.46%	84.12%	As high as possible
Ramachandran Outliers	1.05%	6.98%	6.67%	As low as possible
Bad Bonds	0	0	238	Zero
Bad Angles	24	52	384	Zero

3.5.3.1 Ramachandran validation

A Ramachandran plot interprets phi-psi torsion angles for all residues in the structure. It is a crucial tool for analyzing and validating the geometry and conformations of protein structures, and the behaviour of glycine and proline residues is particularly noteworthy [21] (Table 1).

3.5.3.2 ERRAT validation

The ERRAT program is another important tool used to assess the overall quality and accuracy of protein structure models, complementing the information provided by the Ramachandran plot. The ERRAT program works by calculating an "overall quality factor" for the input protein structure. This quality factor is based on the statistics of non-bonded interactions between different atom types in the structure. The ERRAT score represents the percentage of the protein residues that fall below the 95% confidence limit for reliable no-bonded interactions. Ideally, the majority of residues (80-100%) should fall below the 95% confidence limit, indicating good overall quality. Scores below 70% may suggest issues with the structural model, such as incorrect atom placements, steric clashes, or other problematic non-bonded interactions [22]. In this study, the ERRAT validation outputs for the thurinolysin model showed an overall score of 87.5 % and had the problematic areas appearing in residues around amino acids 70, 100, 140, 370, and 390, as shown in Figure 4. ERRAT indicated that these problematic residues were located far from the active site. As a result, the thurinolysin model successfully passed the validation by ERRAT.

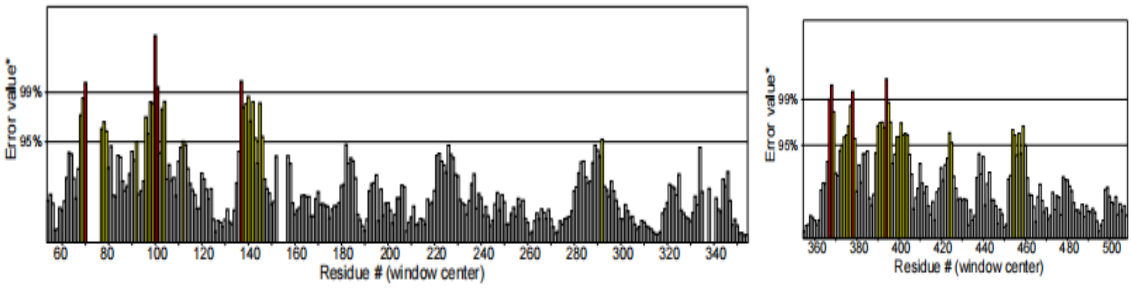

Figure 4: ERRAT validation for the thurinolysin model.

3.5.3.3 Verify 3D validation

The program generates a 3D/1D profile score for each residue, indicating this compatibility. For a high-quality protein structure, at least 80% of the amino acids should have a 3D/1D profile score of 0.1 or greater. Scores below 0, indicating poor compatibility between the 3D structure and the 1D sequence, can point to potential errors in the structural model [23]. The thurinolysin model passed with a value of 81.21% of the residues, with an average 3D-1D score >= 0.1, as illustrated in Figure 5.

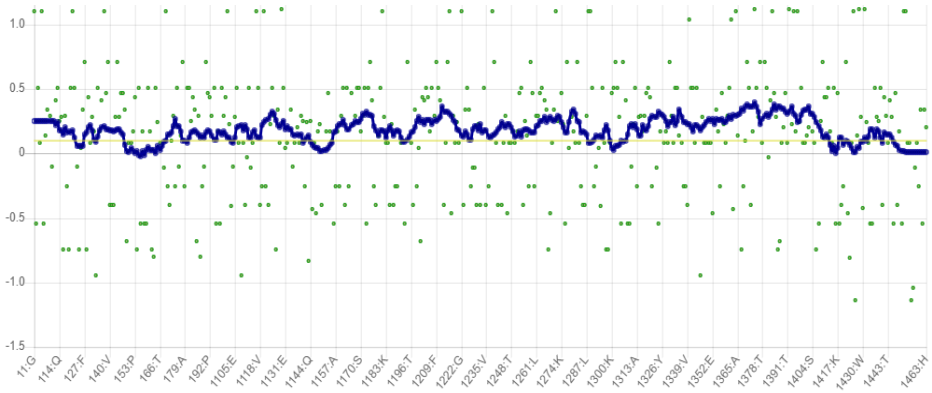

Figure 5: Verify 3D validation for the thurinolysin model

3.5.4 Structural analysis of thurinolysin

The Swiss model of thurinolysin protein, as shown in Figure 6, has overall the same folding as other members of cholesterol-dependent cytolysin, such as anthrolysin O produced by *B. anthracis* (97.53 sequence identity), perfringolysin O produced by *Clostridium perfringens* (72.98 sequence identity), streptolysin O produced by *Streptococcus pyogenes* (63.64 sequence identity), pneumolysin produced by *Streptococcus pneumoniae* (41.42 sequence identity). The structure of the cholesterol-dependent cytolysin member typically consists of four domains, each serving a particular function during the transformation from prepore to pore form (oligomerization state). Domain 1 and domain 2 (D1 and D2) play a role in oligomerisation, domain 3 (D3) is responsible for channel formation by forming β- hairpins, which ultimately insert into the membrane, and domain D4 is responsible for the initial membrane recognition and binding [24].

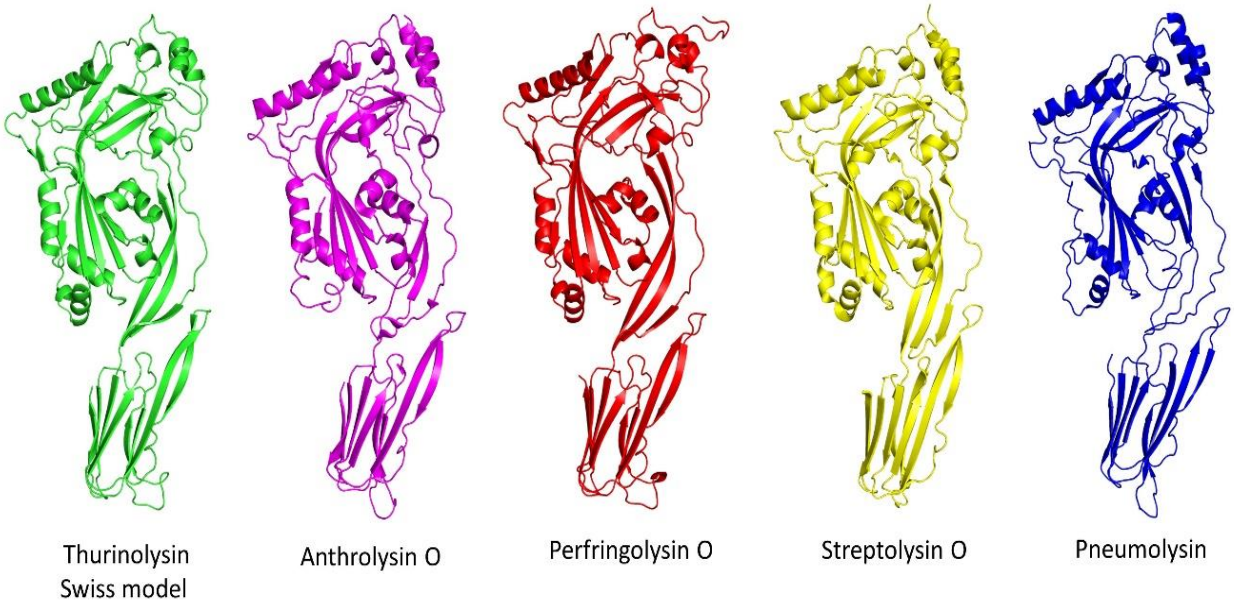

Figure 6: Comparison between the thurinolysin Swiss model with related thiol-activated cytolysin. Anthrolysin O (BDP code: 3cqf), perfringolysin O (BDP code: 1pfo), streptolysin O (BDP code: 4hsc), pneumolysin (BDP code: 4zgh). Cartoon representations of the proteins were generated by PyMOL software

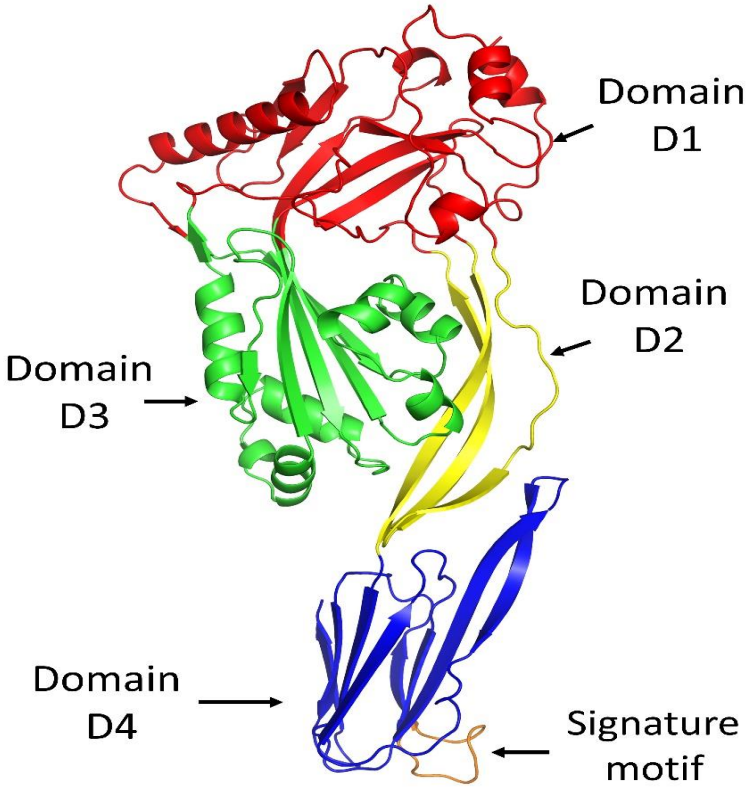

Figure 7: A model structure of thurinolysin generated by Swiss model service coloured by domain. Domain 1 is coloured red, domain 2 is coloured yellow, domain 3 is coloured green, and domain 4 is coloured blue with undecapeptide sequence (ECTGLAWEWWR) coloured orang. A cartoon representation of the model was generated by PyMOL software.

The similarity between the thurinolysin protein model and other cholesterol-dependent cytolysin proteins may provide insight into regions involved in the thurinolysin mechanism of action of thurinolysin. This is due to the similarities in their key functions: D1 and D2 oligomerisation domains, domain D3 β-barrel channel formation, and domain D4 as recognition and binding domain, which contains the signature motif, as depicted in Figure 7.

Despite the similarities between cholesterol-dependent cytolysin proteins in terms of overall structure, domain conformation, and general function. Alterations in the conformation of a signature motif and the orientation of D4 result in functional variances [25]. Subsequent examination of the model revealed that the structural configuration of the signature motif exhibited a greater similarity, precisely an identical resemblance, to that of perfringolysin O compared to other members of the cholesterol-dependent cytolysin family, as demonstrated in Figure 8. This similarity could indicate that they attracted the same target.

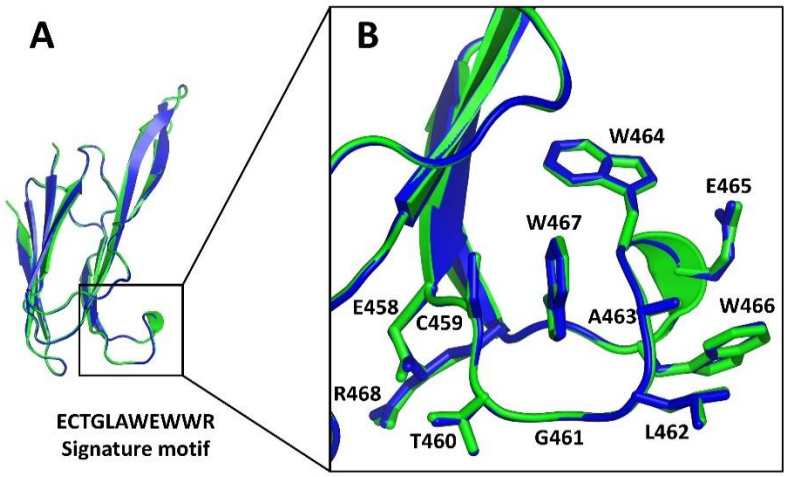

Figure 8: Structural alignment of domain 4. A Structural alignment of domain 4 of thurinolysin Swiss model (blue colour) and perfringolysin O, BDP code: 1pfo (green colour). B Structural alignment of a signature motif (ECTGLAWEWWR).

3.6 Docking of cholesterol

The docking was performed using MOE for the thurinolysin Swiss model. The cholesterol substrate was docked in the Swiss thurinolysin model to see how it interacted with the model and to identify the amino acids that interacted with the substrate, as shown in Figure 9. The docking result for the thurinolysin Swiss model is given in Table 2. Docking was performed, resulting in a binding energy of - 5.25 Kcal/mol, with interactions between 12 amino acid residues and the ligand. In addition, cholesterol was observed to form hydrogen bonds with Gly474 and Thr503 at a distance of 2.29 A°.

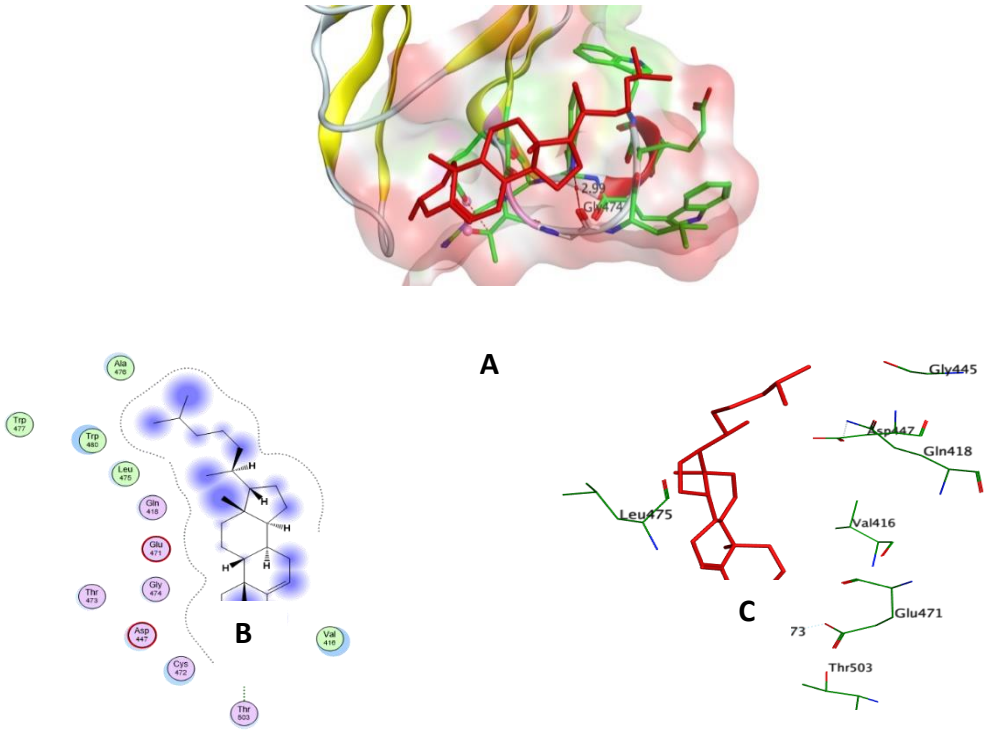

Figure 9: Docking of thurinolysin Swiss mode with cholesterol. (A and B): 3D structure of thurinolysin Swiss mode with cholesterol (in red), compared to undecapeptide sequence (ECTGLAWEWWR)(in green). (C): 2D interaction of docked cholesterol with specific amino acids of thurinolysin Swiss mode.

Table 2: Amino acid residues interact with cholesterol.

Binding energy (Kcal/mol	Hydrophobic interactions	Hydrogen bonds
-5.25	Val 416, Glu,418, Asp447, Glu471, Cys472, Thr473, Leu 475, Ala 476, Trp477 and Trp 480	Gly474 and Thr 503

4. Discussion

Bioinformatics analysis of whole genomic sequences of bacteria advances our knowledge about the distribution of virulence factors and other proteins. A hypothetical protein in Bt whole genomic sequences was subjected to several bioinformatics tools. Based on the results of The NCBI conserved domains and InterPro tools, this hypothetical protein belongs to the pore-forming toxin family, also known as the cholesterol-dependent cytolysin family. These two bioinformatics tools classified bacterial toxins into families based on sequence similarity and domain functions from experimental data. In order to differentiate between various members of thiol-activated cytolysins, they were named based on the microorganism responsible for their production. This nomenclature involved combining the initial portion of the microorganism's name with the suffix "olysin." For instance, *B. anthracis* produces anthrolysin O [26], *Clostridium perfringens* produces perfringolysin O [27], *Streptococcus pyogenes* produces streptolysin O [28], and *Streptococcus pneumonia* produces pneumolysin [29]. The hypothetical protein, produced by the *Bacillus thuringiensis* strain, has been named thurinolysin, using the conventional nomenclature for thiol-activated cytolysin. Another evidence that confirms the thurinolysin protein is a member of thiol-activated cytolysin family was the presence of the signature motif. In molecular biology, a signature motif or sequence motif is an amino-acid sequence pattern that is widespread and usually assumed to be connected to the same biological function of the macromolecule [30].

The signal peptide in thurinolysin was predicted and subsequently removed to obtain a mature protein in order to achieve a precise characterization of the thurinolysin protein. Extracellular toxins are produced with a signal peptide in the cytoplasm of bacteria. As the protein is translocated from the cytoplasm to the environment via any of the secretion systems, the signal peptide is cleaved, resulting in the release of the mature protein into the environment [31]. The physicochemical properties of a protein are essential for studying and characterization of the protein. It is important to calculate an amino acid sequence's theoretical isoelectric point (pI) and molecular weight (Mw), as these values determine the general area of a 2D-gel where a protein of interest may be detected. The extinction coefficient of a protein is known as the amount of light that a protein absorbs at a specific wavelength. When purifying a protein, it is helpful to have an estimate of this coefficient to monitor the protein with a spectrophotometer [32]. The aliphatic index (AI) of thurinolysin was 79.52, indicating the thermally stable of the protein. The AI is known as a positive factor for the increase of thermostability of proteins and is defined as the relative volume occupied by aliphatic side chains (valine, leucine, isoleucine, and alanine). Proteins with AI ranging from 66.5 to 84.33 are known as thermally stable proteins [33].

There will always be some inaccuracies in any molecular model; the number of errors varies according to the target and template's percentage of sequence identity. The target and template have significantly distinct structures if the sequence identity between them is 20% or less [34]. Consequently, the target and template cannot be used for modelling due to the significant number of errors will be considerable, and the model will be inaccurate. Models with a sequence identity ranging from 20% to 90% exhibit a decreasing level of accuracy. In contrast, a sequence identity greater than 90% results in a precise model that, with a few single-side chain exceptions, may be compared to an experimentally determined structure [35]. Therefore,

thurinolysin model is predicted to be an accurate model as the template used in the model building shares a significant sequence identity (95.61).
However, three models were built using different bioinformatics tools that used diverse algorithms for protein structure prediction and building models. As a result, a model with excellent quality needs to be chosen for further research based on model validation. One of the most common validations is the Ramachandran plot validation, it was used to assess the model quality. This validation predicted the structural stereochemical property of the model by measuring the protein's ψ–ϕ main-chain torsion angles, the Ramachandran plot is an approach to identifying energetically preferred regions for backbone dihedral angles in protein structure [21]. MolProbity is a web-based structure-validation service that evaluates the model quality for proteins globally and locally. However, the models' validation is dependent on the intended use. For docking studies with ligands, additional evaluation is necessary, particularly for the active site, and reliable models should have accuracy around the active site.

The basis for molecular modelling is the observation that naturally occurring proteins with identical sequences generally have the same three-dimensional structures and functions [13]. This explains the high similarity between the model of thurinolysin protein and anthrolysin O (the template with 95.61 sequence identity) and other cholesterol-dependent cytolysins. Structural analysis and similarity with cholesterol-dependent cytolysins, especially perfringolysin O, may give insight into the general function of thurinolysin as a pore-forming toxin that undergoes a conformational change from pre-pore to pore form (oligomer) during toxicity, and recognize and targeting the same cells in mammalians.

Conclusions

The integration of multiple bioinformatics methods presents an effective strategy for the investigation of proteins. By using only the DNA sequences of proteins, it is possible to predict and identify various aspects related to protein classification, domain recognition, and physicochemical characteristics. These characteristics include molecular weight, theoretical isoelectric point (PI), extinction coefficient, instability index, and the aliphatic index (AI). These estimations provide preliminary insights into the protein under investigation. The inclusion of such information is crucial in experimental investigations related to proteins. The utilization of protein structural modelling, with a template with a high level of sequence identity, can predict protein functions and link structures to activities. Bt has the capability to synthesize a wide range of toxins that can selectively target both invertebrates and mammalian cells. Moreover, Bt can produce a cholesterol-dependent cytolysin toxin that exhibits both thermostability and a significant degree of sequence identity with anthrolysin O. Furthermore, this toxin exhibits structural similarities with perfringolysin O, which may indicate the similarity in the targeting cells.

Perspectives for Future Research

Produce the thurinolysin protein recombinantly and study its ability to bind with cholesterol and toxicity against mammalian cells, as well as study the anticancer activity of thurinolysin against different human cancer cells.

Author Contributions

All authors contributed to the paper's writing and data analysis. H.S.A. and N.H.O. undertook the design and bioinformatic studies. I.T. carried out docking studies.

Funding

This research received no external funding

Acknowledgements

The authors are thankful to Prof. Dr Colin Berry (School of Biosciences, Cardiff University, Cardiff, UK) for providing the bacterial isolate and whole genome sequence of the bacterium.

Conflict of interests

The authors declare that there is no competing interest.

References

[1] Y. Li, Y. Li, H.M. Mengist, C. Shi, C. Zhang, B. Wang, T. Li, Y. Huang, Y. Xu and T. Jin, "Structural basis of the pore-forming toxin/membrane interaction," *Toxins*, vol. 13, no. 2, p. 128, 2021. doi:10.3390/toxins13020128.

[2] A. Šakanović, N. Kranjc, N. Omersa, S. Aden, A. Kežar, M. Kisovec, A. B. Zavec, S. Caserman, R. J. C. Gilbert, M. Podobnik, A. Crnković and G. Anderluh, "In vitro evolution driven by epistasis reveals alternative cholesterol-specific binding motifs of perfringolysin O," *Journal of Biological Chemistry*, vol. 300, no. 9, pp. 107664–107664, 2024, doi: https://doi.org/10.1016/j.jbc.2024.107664.

[3] B.A. Johnstone, R. Joseph, M.P. Christie, C. J. Morton, C. McGuiness, J. C. Walsh , T. Böcking, R. K. Tweten and M.W. Parker, "Cholesterol-dependent cytolysins: The outstanding questions", *IUBMB Life,* vol. 74, no.12, pp.1-11, 2022, https://doi.org/10.1002/iub.2661.

[4] G. Mendoza-Almanza, E.L. Esparza-Ibarra, J.L. Ayala-Luján, M. Mercado-Reyes, S. Godina-González, M. Hernández-Barrales and J. Olmos-Soto, "The Cytocidal Spectrum of *Bacillus thuringiensis* Toxins: From Insects to Human Cancer Cells", *Toxins,* vol.6, no.12,pp.1-22, 2020, https://doi.org/10.3390/toxins 12050301

[5] M. Gupta, H. Kumar and S. Kaur, "Vegetative Insecticidal Protein (Vip): A Potential Contender From *Bacillus thuringiensis* for Efficient Management of Various Detrimental Agricultural Pests", *Frontiers in Microbiology*, vol.13, no.12, pp.1-28, 2021, doi:10.3389/fmicb.2021.659736.

[6] L. Palma, D. Muñoz, C. Berry, J. Murillo, and P. Caballero, "*Bacillus thuringiensis* Toxins: An Overview of Their Biocidal Activity," *Toxins*, vol. 6, no. 12, pp. 3296–3325, 2014, doi: https://doi.org/10.3390/toxins6123296.

[7] K. Li, M. Chen, J. Shi, and T. Mao, "An overview of the production and use of *Bacillus thuringiensis* toxin," *Open Life Sciences*, vol. 19, no. 1, 2024, doi: https://doi.org/10.1515/biol-2022-0902.

[8] X. Qiu, H. Li, G. Ver Steeg, and A. Godzik, "Advances in AI for Protein Structure Prediction: Implications for Cancer Drug Discovery and Development," *Biomolecules*, vol. 14, no. 3, p. 339, 2024, doi: https://doi.org/10.3390/biom14030339.

[9] S. Kaczanowski and P. Zielenkiewicz, "Why similar protein sequences encode similar three-dimensional structures?," Theoretical Chemistry Accounts, vol. 125, no. 3–6, pp. 643–650, 2009, doi: https://doi.org/10.1007/s00214-009-0656-3.

[10] S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman, "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Research*, vol. 25, no. 17, pp. 3389–3402, 1997, doi: https://doi.org/10.1093/nar/25.17.3389.

[11] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R.D. Appel and A. Bairoch, "ExPASy: The proteomics server for in-depth protein knowledge and analysis," *Nucleic Acids Research*. Vol. 31, no.13, pp. 3784-3788,2003, doi:10.1093/nar/gkg563.

[12] F. Kiefer, K. Arnold, M. Kunzli, L. Bordoli, and T. Schwede, "The SWISS-MODEL Repository and associated resources," *Nucleic Acids Research*, vol. 37, pp. 387–392, 2009, doi: https://doi.org/10.1093/nar/gkn750.

[13] Y. Zhang, "I-TASSER server for protein 3D structure prediction," *BMC Bioinformatics*, vol. 9, no. 1, 2008, doi: https://doi.org/10.1186/1471-2105-9-40.

[14] L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, and M. J. E. Sternberg, "The Phyre2 Web Portal for Protein modeling, Prediction and Analysis," *Nature Protocols*, vol. 10, no. 6, pp. 845–858, 2015, doi: https://doi.org/10.1038/nprot.2015.053.

[15] G. M. Morris, R. Huey, W. Lindstrom, M. F Sanner, R. K. Belew, D. S. Goodsell, and A. J. Olson, "AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility,"

Journal of computational chemistry, vol. 30, no.16,pp.2785–2791,2009 https://doi.org/10.1002/jcc.21256.

[16] J. J. Armenteros, K.D. Tsirigos, C.K. Sønderby, T.N. Petersen, O. Winther, S. Brunak, G. von Heijne and H. Nielsen, "SignalP 5.0 improves signal peptide predictions using deep neural networks," *Nature Biotechnology*, vol. 37, no. 4, pp. 420–423, 2019, doi: https://doi.org/10.1038/s41587-019-0036-z.

[17] D. G. Gamage, A. Gunaratne, G. R. Periyannan, and T. G. Russell, "Applicability of Instability Index for In vitro Protein Stability Prediction," *Protein and Peptide Letters*, vol. 26, no. 5, pp. 339–347, 2019, doi: https://doi.org/10.2174/0929866526666190228144219.

[18] T. Hameduh, Y. Haddad, V. Adam, and Z. Heger, "Homology Modeling in the Time of Collective and Artificial Intelligence," *Computational and Structural Biotechnology Journal*, 2020, doi: https://doi.org/10.1016/j.csbj.2020.11.007.

[19] L. Chen, Q. Li, K.F.A. Nasif, Y. Xie, B. Deng, S. Niu, S. Pouriyeh, Z. Dai, J. Chen and C.Y. Xie, "AI-Driven Deep Learning Techniques in Protein Structure Prediction," *International Journal of Molecular Sciences*.vol.25, no.15, pp. 1-21,2024, https://doi.org/10.3390/ijms25158426.

[20] I. M. Taban, J. Zhu, H. F. DeLuca, and C. Simons, "Synthesis, molecular modelling and CYP24A1 inhibitory activity of novel of (E)- N -(2-(1 H -imidazol-1-yl)-2-(phenylethyl)-3/4-styrylbenzamides," *Bioorganic and Medicinal Chemistry*, vol. 25, no. 15, pp. 4076–4087, 2017, doi: https://doi.org/10.1016/j.bmc.2017.05.055.

[21] S. W. Park, B. H. Lee, S. H. Song, and M. K. Kim, "Revisiting the Ramachandran plot based on statistical analysis of static and dynamic characteristics of protein structures," *Journal of Structural Biology*, vol. 215, no. 1, p. 107939, 2023, doi: https://doi.org/10.1016/j.jsb.2023.107939.

[22] F. A. Binbay, D. C. Rathod, P. George, and D. Imhof, "Quality Assessment of Selected Protein Structures Derived from Homology Modeling and AlphaFold," *Pharmaceuticals*, vol. 16, no. 12, pp. 1662–1662, 2023, doi: https://doi.org/10.3390/ph16121662.

[23] J. Aguirre-Plans, A. Meseguer, R. Molina-Fernandez, M. A. Marín-López, G. Jumde, K. Casanova, J. Bonet, O. Fornes, N. Fernandez-Fuentes and B. Oliva, "SPServer: split-statistical potentials for the analysis of protein structures and protein–protein interactions," *BMC Bioinformatics*, vol. 22, no. 1, 2021, doi: https://doi.org/10.1186/s12859-020-03770-5.

[24] J.C. Evans, B.A. Johnstone, S.L. Lawrence, C.J. Morton, M.P. Christie, M.W. Parker and R.K. Tweten. "A Key Motif in the Cholesterol-Dependent Cytolysins Reveals a Large Family of Related Proteins," *American Society for Microbiology*, vol. 11, no. 5, 2020, doi: https://doi.org/10.1128/mbio.02351-20.

[25] H. Jeong, K. Grimes, K. K. Rauwolf, P. Bruch, T. Rausch, P. Hasenfeld, E. Benito, T. Roider, R. Sabarinathan, D. Porubsky, S. A. Herbst, B. Erarslan-Uysal, J. Jann, T. Marschall, D. Nowak, J.Bourquin, A. E. Kulozik, S. Dietrich, B. Bornhauser, A.D. Sanders and J. O. Korbel, "Functional analysis of structural variants in single cells using Strand-seq," *Nature Biotechnology*, vol. 41, no. 6, pp. 832–844, 2022, doi: https://doi.org/10.1038/s41587-022-01551-4.

[26] K. A. Johnson and A. Radhakrishnan, "The use of anthrolysin O and ostreolysin A to study cholesterol in cell membranes," *Methods in enzymology*, vol. 649, pp. 543–566, 2021, https://doi.org/10.1016/bs.mie.2021.01.011.

[27] A. Camargo, Juan David Ramírez, R. Kiu, L. J. Hall, and M. Muñoz, "Unveiling the pathogenic mechanisms of *Clostridium perfringens* toxins and virulence factors," *Emerging Microbes & Infections*, vol. 13, no. 1, 2024, doi: https://doi.org/10.1080/22221751.2024.2341968.

[28] J. Clarke, M. Baltazar, M. Alsahag, S. Panagiotou, M. Pouget, W. A. Paxton, G. Pollakis, D. Everett, N. French and A. Kadioglu, "Streptolysin O concentration and activity is central to in vivo phenotype and disease outcome in Group A Streptococcus infection," *Scientific Reports*, vol. 11, no. 1, p. 19011, 2021, doi: https://doi.org/10.1038/s41598-021-97866-4.

[29] J. M. Pereira, S. Xu, J. M. Leong, and S. Sousa, "The Yin and Yang of Pneumolysin During Pneumococcal Infection," *Frontiers in Immunology*, vol. 13, 2022, doi: https://doi.org/10.3389/fimmu.2022.878244.

[30] F. Silva, I. Boal-Carvalho, N. Williams, M. Chabert, C. Niu, D. Hedhili, H. Choltus, N. Liaudet, N. Gaïa, W. Karenovics, P. Francois and M. Schmolke, "Identification of a short sequence motif

in the influenza A virus pathogenicity factor PB1-F2 required for inhibition of human NLRP3," *Journal of Virology*, vol. 98, no. 5, 2024, doi: https://doi.org/10.1128/jvi.00411-24.

[31] S. Kaushik, H. He, and R. E. Dalbey, "Bacterial Signal Peptides- Navigating the Journey of Proteins," *Frontiers in Physiology*, vol. 13, 2022, doi: https://doi.org/10.3389/fphys.2022.933153.

[32] M. F. Pignataro, M. G. Herrera, and V. I. Dodero, "Evaluation of Peptide/Protein Self-Assembly and Aggregation by Spectroscopic Methods," *Molecules*, vol. 25, no. 20, p. 4854, 2020, doi: https://doi.org/10.3390/molecules25204854.

[33] A. T. Nishimoto, J. W. Rosch, and E. I. Tuomanen, "Pneumolysin: Pathogenesis and Therapeutic Target," *Frontiers in Microbiology*, vol. 11, 2020, doi: https://doi.org/10.3389/fmicb.2020.01543.

[34] B. J. Bender, B. Marlow, and J. Meiler, "Improving homology modeling from low-sequence identity templates in Rosetta: A case study in GPCRs," *PLOS Computational Biology*, vol. 16, no. 10, p. e1007597, 2020, doi: https://doi.org/10.1371/journal.pcbi.1007597.

[35] R. D. Ahmed, H. S. Auhim, H. L. Worthy, and D. D. Jones. "Fluorescent Proteins: Crystallization, Structural Determination, and Nonnatural Amino Acid Incorporation," *Methods in Molecular Biology* (Clifton, NJ),vol. 2564,p. 99-119. 2023.