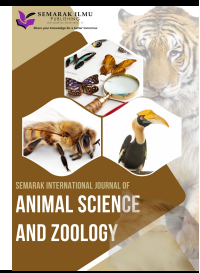




Semarak International Journal of Animal Science and Zoology

Journal homepage:
<https://semarakilmu.my/index.php/sijaszi/index>
ISSN: 3083-9254



Spider Silk Unravelling: Structural Insights into Ampullate Spidroin via In Silico Modeling from the Cellar Spider, *Crossopriza lyoni*

Johan Ariff Mohtar^{1,3,*}, Mohd Faidz Mohamad Shahimin^{1,2,3}, Amirul Ridzuan Abu Bakar^{1,3}, Khadijah Hanim Abdul Rahman¹

¹ Faculty of Chemical Engineering & Technology, Universiti Malaysia Perlis, 02100 Padang Besar, Perlis, Malaysia

² Centre of Excellence Water Research and Environmental Sustainability Growth (WAREG), Kompleks Pusat Pengajian Jejawi 3, 02600 Arau, Perlis, Malaysia

³ Bioresource and Food Special Interest Group, Faculty of Chemical Engineering & Technology, Universiti Malaysia Perlis, 02600 Arau Perlis, Malaysia

ARTICLE INFO

Article history:

Received 30 July 2025

Received in revised form 8 September 2025

Accepted 10 October 2025

Available online 28 October 2025

Keywords:

Dragline silk; *In silico*; modeling; NT-AmSp; Spidroin; structure

ABSTRACT

Spider dragline silk, or major ampullate silk, is a remarkably high-performing polymeric biomaterial with outstanding physical and mechanical properties due to its silk proteins, called spidroins. In particular, the N-terminus of spidroins plays a crucial role in silk fibre formation. Extensive research on the 3D N-terminal structure of major ampullate spidroins (MaSps) in orbicularian spiders has led to the neglect of major ampullate silks from non-orbicularians, despite their unique silk protein sequence, with a predominant focus on the adult stage rather than the early developmental stage. In this study, we elucidated the N-terminal ampullate spidroin (NT-AmSp) structure from the prenymp of the non-orbicularian species, *Crossopriza lyoni*. The NT-AmSp sequence of 155 amino acids was subjected to protein homology modeling, threading, and ab initio modeling through multiserver-based in silico predictions using SWISS-MODEL, Phyre2, and I-TASSER, respectively. The quality of each generated model was analysed using ProSA-web, QMEAN, and SAVES (parameters i.e., ERRAT, Verify3D, and Ramachandran plot) servers. Finally, the models were superimposed with an NMR-determined NT-MaSp from *E. australis* for similarity assessment using SuperPose. Models ranked first by both SWISS-MODEL and Phyre2 (Model 1) and Model 3 from I-TASSER with the highest C-score were chosen as the best predicted models. All models possessed five α -helices except for Model 3 with an additional α -helical conformation representing the signal peptide region. Overall, the models were of relatively good quality according to the analysis. The structure superimposition with *E. australis* NT-MaSp1 (4FBS) yielded an acceptable RMSD value between 2.0 Å and 3.0 Å. In silico structural modeling proves to be a powerful tool for assessing protein molecular functions. Elucidating the N-terminal structure of AmSp from *C. lyoni* prenymp may enhance our understanding of spidroin N-terminal diversity across spider taxa between adult and nymphal stages, and aid in developing soluble tags for heterologous protein expression.

* Corresponding author.

E-mail address: johanariff@unimap.edu.my

1. Introduction

Spider silk is a polymeric biomaterial that demonstrates significantly superior performance compared to silkworm-derived fibre [1]. A single spider species may have the capacity to produce up to eight types of silks. Among these, dragline silk, also referred to as major ampullate silk has emerged as a highly promising candidate with its outstanding physico-mechanical properties. Dragline silk is utilized as a safety line during crawling or as the structural framework of a web. Its toughness and extensibility exceed that of the best synthetic high-performance fibre, Kevlar [2,3], and is even five times stronger than steel on a weight-by-weight basis [4]. The unique attributes of this silk are primarily attributed to the presence of the silk protein known as spidroin, underscoring its potential significance in the development of innovative and advanced biocomposite materials.

Similar to other silk types, a dragline silk thread is packed into two layers—the outer shell and the fibrillar core—similar to silkworm fibre [5]. It is composed of polypeptide stretches of spidroin that primarily constitute more than 90% of the fibre core [6]. Each type of spider silk is composed of a combination of at least two spidroin types that tailor its properties and functions [7]. The physico-mechanical properties of a dragline silk are influenced by the ratio of two major ampullate spidroins, MaSp1 and MaSp2 which in general, varies among different species across spider taxa [8]. Spidroin is a naturally occurring scleroprotein with a substantial molecular mass of up to ca. 660 kDa in its native state [9]. They are produced in specialized silk glands such as major ampullate and minor ampullate glands located in the abdomen, and stored as highly concentrated liquid crystalline granules (dope). The dope undergoes thermodynamic changes as it flows through the spinning duct during fibre extrusion [10]. In general, the structural organization of spidroin is characterized by three domain sequences, with a large proportion of an alternating repetitive core domain in the middle, flanked by the N- and C-terminal sequences. The repetitive core domain plays a major role in conferring physico-mechanical properties, such as toughness, strength, elasticity, and supercontraction, to silk fibre, whereas the terminal domains are primarily involved in fibrillar formation [11].

The N-terminal domain of a spidroin is approximately 150 amino acid region that exhibits pH-dependent behavior, playing a crucial role in regulating spidroin solubility during storage and fibrillar assembly [11]. At neutral pH, this domain adopts a monomeric five-helix bundle structure but under acidic conditions, it undergoes dimerization, hindering aggregation and influencing solubility at higher concentrations [12]. The pH-dependent transitions in this domain are particularly relevant for its functions during storage and assembly, ensuring optimal spidroin behaviour within a specific pH range [12]. Like many secretory proteins, the N-terminus also controls the secretion of spidroins from silk glands due to the presence of a signal peptide region. The signal peptide is typically located in the extreme N-terminal region, right after the translational start site, which contains the first methionine that adopts an α -helical conformation [13]. The N-terminus of spidroin is further characterized by an unusual abundance of methionine residues in the hydrophobic core, contributing to the protein's ability to dynamically change shape and optimize its function, as elucidated by Heiby *et al.*, [14]. Additionally, beyond the common features of spidroin, a non-coding sequence upstream of the start codon was identified to contain putative transcription promoter motifs, including CACG (300 bp) and the classic eukaryotic promoter motif TATA (150 bp), as reported by Chaw *et al.*, [15].

Over the years, spider dragline silk has gained recognition as an exceptionally promising contender for various potential biocomposite applications. The exploration of spider dragline silk in this context opens up new avenues for leveraging its strength, flexibility, and other remarkable characteristics in diverse applications within the field of biocomposite, showcasing its potential to revolutionize and enhance various industries. In recent years, innovations in spider silk technology

have primarily focused on manipulating dragline silk from araneoid-deinopid (orbicularian) adult spiders for synthetic silk production, owing to their high performance and outstanding properties compared to non-orbicularians, such as *Crossopriza lyoni*. Orbicularians have the ability to construct orb-type webs. As a result, extensive research has been conducted on the elucidation of three-dimensional (3D) structure of N-terminal domain of major ampullate spidroins from various orbicularian species, investigating their structure-function relationship for the production of recombinant MaSps in heterologous hosts to synthesize artificial spider silk [16]. This emphasis has led to the neglect of major ampullate silks from non-orbicularians, as they were initially believed to possess fewer desirable properties [17]. Nevertheless, recent studies have revealed that non-orbicularian originated silk gene sequences are frequently distinct in their sequence arrangement, presenting unique characteristics [18]. For example, spidroin from the Syspermiata subgroup exhibits a distinctive molecular motif pattern in the repetitive region [18]. This signals the potential diversity of spidroin sequences across spider taxa. To date, few or no studies have been conducted on the structural prediction of the N-terminal domain of MaSps from any non-orbicularian species.

C. lyoni is a non-orbicularian cellar spider with no economic importance and poses no harm to humans. Often considered beneficial for pest control, especially against mosquitoes and flies, this species has long dominated human residences in Malaysia. When their numbers are high, they can become quite pestiferous due to their extensive webbing. Until now, MaSps in *C. lyoni* have not been extensively documented, with only one reported study by [19], highlighting the MaSp amino acid sequences from mature adults in Malaysia and India. In the present study, the N-terminal MaSp sequence from prenymp *C. lyoni* was examined. However, since the characteristics of the spidroin could not be homologized with that of orbicularians, it is generally regarded as ampullate spidroin (AmSp) “(Mohtar), unpublished data”. Spiders exhibit spinning behavior as early as after egg hatching, but no studies have delved into characterizing the structure of ampullate spidroin from the prenympal stage. Thus, as an initiative effort, the present study aims to elucidate the predictive protein structure of the N-terminal domain of AmSp from *C. lyoni* prenymps.

2. Methodology

2.1 Protein Sequence Retrieval

The N-terminal domain of ampullate spidroin of *C. lyoni* (GenBank accession number: WBU86978.1), designated as NT-AmSp was used in the prediction of the protein structure (Figure 1). The 207 amino acid sequence was previously characterized from the transcriptomes of the whole body of *C. lyoni* prenymp “(Mohtar), unpublished data”. The NT-AmSp (WBU86978.1) consists of two regions, the N-terminus (Met1-Tyr155) where a putative signal peptide was located between Met1-Ala19 and the partial repetitive domain (Gly156-Gly207) at the 3' end of the sequence (Figure 1). Prior to modeling, the repetitive region was removed yielding a refined NT-AmSp of 155 amino acid (aa) residues.

```
>WBU86978.1 ampullate spidroin variant 1, partial [Crossopriza lyoni]
MGWFGAIVVLPLLWNAVIASSSSTLWSNPSLALGFISNLGAYLGNISGVFTGDQISDINDISASISGSINA
YASSGRQSPAMLKALDMAFASSLAEIVASEGGLNIAAQTSAITNSMSQAFASVGLPVDKQFIYEINELIT
MFANDALAAASNEVAYGGGSASAAAAAAGGGGGYGGYGAGAGAAAAAAGAGGAGQGGYGQAGAG
```

Fig. 1. The amino acid sequence of NT-AmSp from *C. lyoni* prenymp, revealing two distinct regions. The N-terminus, highlighted in cream, includes a putative signal sequence (underlined) at the beginning of the sequence, and the partial repeat domain is marked in green

2.2 Protein Structure Prediction

The *in silico* prediction of NT-AmSp protein structure (155 aa) was performed in a monomeric state by multiserver approach using three prediction platforms, i.e., SWISS-MODEL (<https://swissmodel.expasy.org/>) [20], Phyre2 (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>) [21] and I-TASSER (<https://zhanggroup.org/I-TASSER/>) [22-23]. SWISS-MODEL facilitates homology modeling by aligning the amino acid sequence of the target protein to known homologous protein templates. While BLAST is commonly used for sequence alignment, SWISS-MODEL employs a combination of sequence and structure-based methods. The number of generated models depends on user preferences, and multiple models can be produced [20]. Phyre2 predicts a target protein structure through threading with known template proteins using PSI-BLAST based on amino acid sequences in the Protein Data Bank (PDB). Following secondary structure prediction by PSIPRED, it starts to build hidden Markov (HMM) model of sequence prior to threading in HMM database of known structures. It also combines both homology modeling and *ab initio* methods, generating one top model by default, although additional models can still be accessed. In cases where a homologous protein template is unavailable, Phyre2 incorporates *ab initio* or *de novo* folding methods [21]. I-TASSER performs protein threading from the PDB library by LOMETS algorithm and then reassemble the excised template fragments into full-length by Monte Carlo simulations. When no appropriate template is identified, it integrates *ab initio* modeling methods to predict, refine and improve the model. It ranks multiple structural models and generates five top structures based on the C-score [23]. All *in silico* predictions were performed at default settings on each server throughout the experiment.

2.3 Quality and Similarity Assessments

The quality of each generated protein model was analyzed using three web-based servers: Protein Structure Analysis (ProSa-web) (<https://prosa.services.came.sbg.ac.at/prosa.php>) [24], Qualitative Model Energy Analysis (QMEAN) version 4.3.0 (<https://swissmodel.expasy.org/qmean/>) [25], and Structural Analysis and Verification Server (SAVES) version 6.0 (<https://saves.mbi.ucla.edu/>) [26]. The similarity assessment of the three predicted models of NT-AmSp was conducted using SuperPose version 1.0 (<http://superpose.wishartlab.com/>) [27]. In brief, the experimentally determined 3D structure of monomeric NT-MaSp1 protein (ID: 4FBS) was retrieved from the Protein Data Bank (PDB) (<https://www.rcsb.org>) and utilized for structural superimposition. The NT-MaSp1 protein from the nurse web spider, *Euprosthenops australis*, resolved through X-Ray diffraction method [28], encompasses a complete N-terminal domain of major ampullate spidroin, represented by only one chain (A) with five helical structures across the amino acid sequence. The superimposed protein complexes were quantified using the root-mean-square deviation (RMSD) value, and the structures were visualized with MolScript integrated into the server.

3. Results

3.1 *In silico* Structural Prediction

The three *in silico* prediction servers successfully generated a total of 11 monomeric form of NT-AmSp models: 5 by SWISS-MODEL, 1 by Phyre2, and 5 by I-TASSER. The best models were selected based on the highest ranks, with Model 1 from SWISS-MODEL and Phyre2, and Model 3 from I-TASSER, which had the highest C-score of -0.80, being chosen. In I-TASSER, the C-score serves as a

confidence measure, calculated from the significance of threading template alignments and convergence parameters of structure assembly simulations. The typical C-score ranges between -5 to 2, where a higher value indicates better model quality. All models from SWISS-MODEL and Phyre2 were predicted within a similar amino acid range, spanning from 21 to 155, as detailed in Table 1, representing the entire N-terminus domain.

Table 1

Amino acid positions in N-terminal domain of AmSp for each predicted model

Server	Model	Template	Amino acid range	Region
SWISS-MODEL	1	7wio.1.A	29 - 146	NT
Phyre2	1	c3lr6A	22 - 145	NT
I-TASSER	3	8gs7A, 7wioA, 8gs7, 3lr2, 2n3eA, 8gs7-1, 3lr8A, 7aOiA, 7wioA-1, 2lpiA	3 - 148	SP-NT

NT = N-terminus, SP = signal peptide

The 3D protein structure consists of five α -helical bundles across the domain (Figure 2A, 2C and 2E). Both SWISS-MODEL and Phyre2 models heavily depend on the size of homologous MaSp templates in the PDB, which predominantly encompass the N-terminal domain. Thus, the prediction occurred in an almost identical amino acid region. However, I-TASSER, utilizing an *ab initio* method, observed an additional α -helical conformation at the extreme end between residues 3 to 19 in Model 3 (Figure 3). This region was previously annotated as the putative signal peptide region of the N-terminus “(Mohtar), unpublished data”.

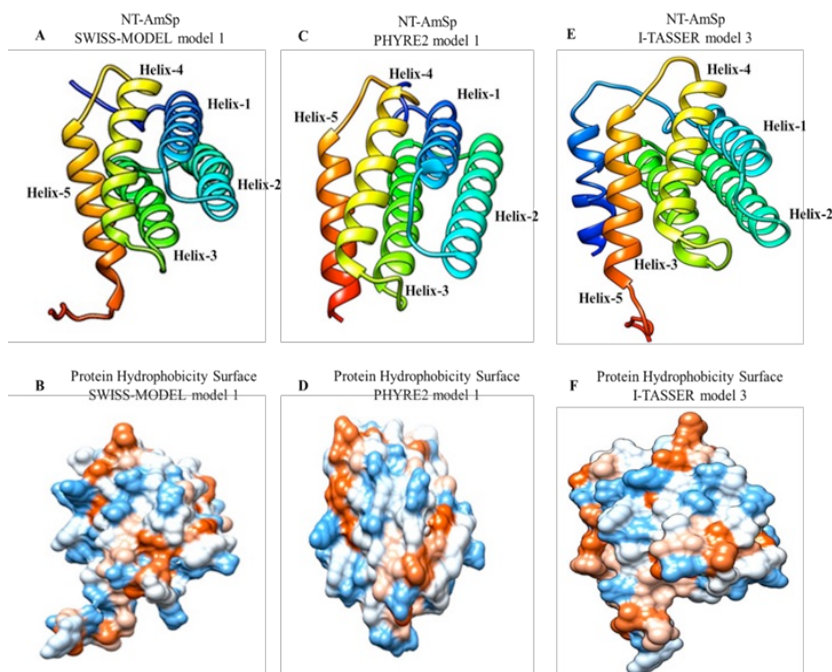


Fig. 2. The selected best models of tertiary structures of NT-AmSp monomer predicted by (A) SWISS-MODEL (Model 1) and its (B) hydrophobicity surface, (C) Phyre2 (Model 1) and its (D) hydrophobicity surface, and (E) I-TASSER (Model 3) with its (F) hydrophobicity surface. The surface representation displays exposed charged residues, with acidic shown in red and basic in blue

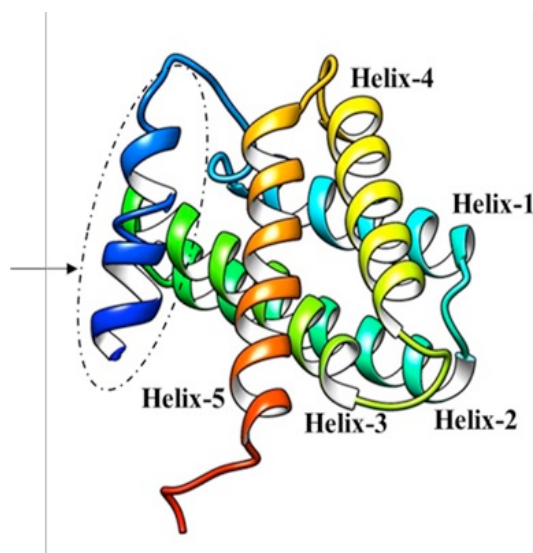


Fig. 3. The predicted Model 3 of the NT-AmSp protein structure by I-TASSER, highlighting an additional α -helical conformation between Trp3 to Ala19 (enclosed in a dashed circle with an arrow), representing the signal peptide region of the domain

3.2 Structural Quality Assessment

The quality of the three selected models was assessed using interactive web servers, namely ProSA-web, QMEAN, and SAVES version 6.0 (Table 2). The evaluation by ProSA-web and QMEAN was based on the z-score and QMEAN4 score, respectively. Meanwhile, SAVES version 6.0 assessed the models using the percentage of ERRAT, Verify3D, residues in the allowed region, and the overall G-factor. The latter two parameters were obtained from the Ramachandran plot via PROCHECK program version 3.5. As shown in Table 2, both protein models from SWISS-MODEL and Phyre2 displayed z-scores of -6.47 and -6.91, respectively, falling within the range of experimentally-determined protein chains.

Table 2

The quality of NT-AmSp predicted models by ProSA-web, QMEAN and SAVES

Server	Model	ProSA z-score	QMEAN4 Score	SAVES v6.0			Total residues in all allowed region (%)
				ERRAT (%)	Verify3D (%)	Overall G- factor	
SWISS-MODEL	1	-6.47	-3.46	97.60	48.15	0.08	99.1
Phyre2	1	-6.91	-2.26	92.24	54.84	0.3	100.0
I-TASSER	3	-5.50	-4.70	87.76	49.03	-0.28	100.0

Threshold value of good quality model: ProSA Z-score = -10 to 10, QMEAN4 score = higher score signifies better quality, ERRAT (%) = residues > 50%, Verify3D (%) = residues $\geq$ 80%, overall G-factor = 0 to -0.5 and total residues in all allowed region (%) = > 90%. ND = not determined.

In ProSA-web, the z-score indicates the overall model quality and is plotted against the z-scores of all experimentally determined protein structures (X-ray crystallography and NMR spectroscopy) in the current PDB. It measures the deviation of the total energy of the protein structure [29]. A z-score within the range of scores found for native proteins of similar size suggests good model quality, while a score outside the range signifies a potentially erroneous structure [24]. The acceptable range for the z-score of native proteins is between -10 to 10, within the space of proteins related to X-ray [30]. Hence, Model 1 by both SWISS-MODEL and Phyre2 and I-TASSER Model 3 are considered to be of good quality with z-scores of -6.47, -6.91 and -5.50, respectively.

Meanwhile, the QMEAN4 scores of Model 1 by Phyre2 recorded the highest value at -2.26, followed by those of SWISS-MODEL and I-TASSER, each at -3.46 and -4.70, respectively. The score measures the overall quality of the protein structure based on geometric values, non-covalent interactions, and consistency of the structure [31]. It indicates that the higher the score, the better the quality of the protein structure [31]. Thus, Model 1 by Phyre2 exhibits better quality, followed by Model 1 by SWISS-MODEL and Model 3 of I-TASSER, accordingly. Further assessments of the three selected protein models in SAVES v6.0 indicated that their overall quality, as evaluated by ERRAT, was high. Model 1 of both SWISS-MODEL and Phyre2 recorded scores of 97.60% and 92.24%, respectively, while Model 3 by I-TASSER scored 87.76%. ERRAT measures the overall quality factor for non-bonded atomic interactions, where higher scores indicate higher quality. Typically, a score of more than 50% is accepted as indicative of a high-quality model [32].

Verify3D calculates the compatibility of a 3D atomic model with its own 1D amino acid sequence to assess the three-dimensional structure. A value higher than 80% generally indicates good model quality [33]. Nevertheless, all models generated in this study possessed low Verify3D values, indicating that fewer than 80% of the amino acid residues had an average 3D-1D score ≥ 0.1 . Despite these lower values, it does not necessarily imply that the models are of low quality. This observation may be associated with the size of NT-AmSp, which has a length of less than 160 amino acids. While there is no strict minimum length requirement for a protein in Verify3D, a protein of reasonable length with sufficient structural complexity is highly recommended. In practice, proteins with fewer than 50–100 residues may not provide enough structural information for a robust Verify3D analysis, possibly explaining the lower Verify3D values observed for the models. Hence, the use of Verify3D in the present study may not be compatible for assessing the 3D structure of NT-AmSp models.

From the Ramachandran plots, the G-factor values for the selected models were computed as 0.08, 0.3, and -0.28 for SWISS-MODEL, Phyre2, and I-TASSER models, respectively (Table 2). The G-factor assesses main chain bond lengths and angles, and acceptable values typically range from 0 to -0.5, with higher quality models displaying values close to or slightly below 0 [34]. All NT-AmSp models, therefore, indicated good quality. In terms of the percentage of residues in the allowed region, the Ramachandran plot revealed that 95.8% of amino acids in SWISS-MODEL Model 1 resided in the core favorable region, while the remaining 3.3% were in the additional allowed region (Figure 4A). No residues were observed in the generally allowed region, with less than 1% of residues in the disallowed region. The quality of a protein structure is generally considered good when the fraction of non-glycine residues in the outlier region is $< 15\%$, with smaller fractions indicating better model quality [35]. Considering this criterion, SWISS-MODEL Model 1 is still regarded as of good quality, despite the observed small residue fraction in the disallowed region. For Phyre2 Model 1, 94.5% of residues were in the most favorable region, followed by 4.6% and 0.9% in the additional and generously allowed regions (Figure 4B). Meanwhile, I-TASSER Model 3 showed 87.6% of residues in the most favorable region, with the remaining 10.2% and 2.2% in the additional and generously allowed regions (Figure 4C). Predicted protein models with over 90% of residues in the allowed region are considered of good quality [33]. Hence, the predicted NT-AmSp structures are deemed a good

model, with 99.1%, 100.0%, and 100.0% of the total residues present in the all allowed region of the Ramachandran plot as tabulated in Table 2.

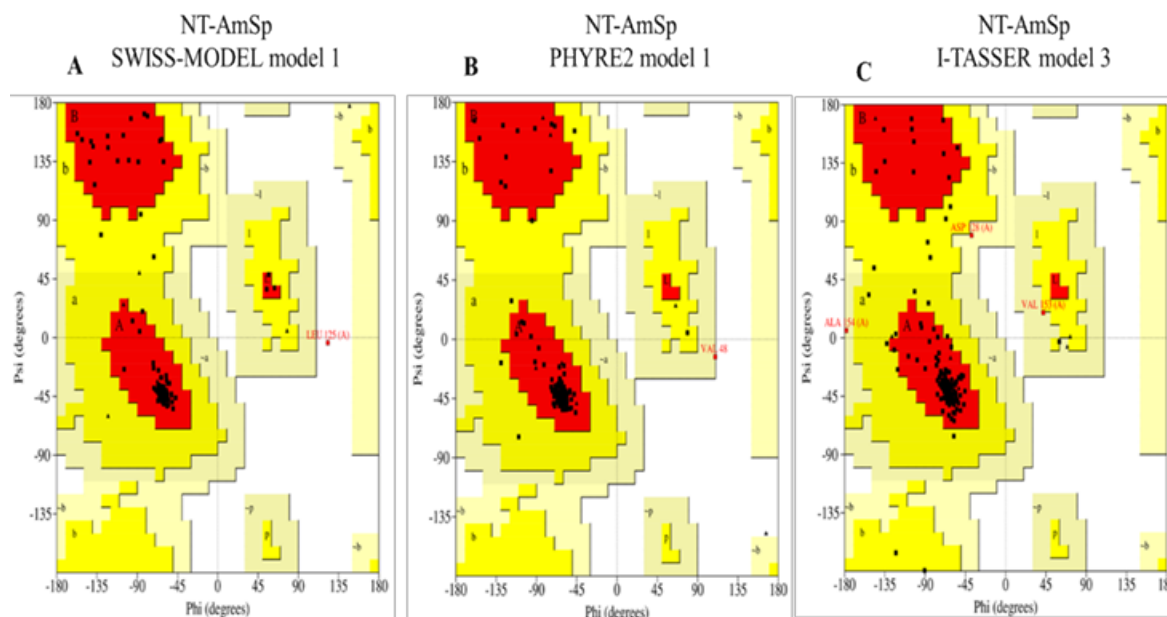


Fig. 4. The Ramachandran plots depicting the backbone dihedral angles of three NT-AmSp models generated by different modeling tools: (A) SWISS-MODEL, (B) Phyre2, and (C) I-TASSER. The plots were obtained through PROCHECK analysis on SAVES version 6.0. Over 90% of the total residues in each model were found within the all allowed regions of the Ramachandran plot, indicating favorable backbone conformations

3.3 Structural Similarity Assessment

The structural similarity between the selected NT-AmSp models and the experimentally determined structure of *E. australis* NT-MaSp by X-ray diffraction (Figure 5A) was analyzed using SuperPose version 1.0.

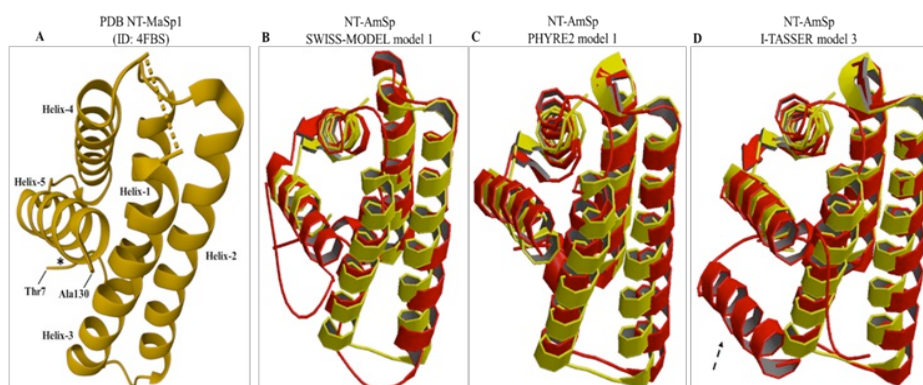


Fig. 5. Structural superimposition comparing (A) the experimentally determined N-terminal domain of major ampullate spidroin (4FBS) from the nursery web spider *E. australis* with the best-predicted NT-AmSp models generated by (B) SWISS-MODEL, (C) Phyre2, and (D) I-TASSER. The crystal models revealed five helical structures (shown in red) that were reasonably well superimposed with those of NT-MaSp1 (4FBS) in yellow, with labeled first and last protein residues. The additional helix indicated by the dashed arrow in (D) did not superimpose with the corresponding region in the PDB NT-MaSp1 structure. An asterisk symbolizes the unmodeled loops between Thr7-Thr11 in this region (A)

Superimposition yielded RMSD scores of 2.17, 2.49, and 2.04 for SWISS-MODEL Model 1, Phyre2 Model 1, and I-TASSER Model 3, respectively (Table 3). In the context of comparing different protein structures, an acceptable range for RMSD values typically falls between 2.0 Å to 3.0 Å, where RMSD $\leq$ 2.0 Å is considered highly similar and $\geq$ 3.0 Å indicates less similarity [36]. The relatively low RMSD values (2.17, 2.49, and 2.04) suggest a high degree of similarity between the 3D structures of *C. lyoni* NT-AmSp models and *E. australis* NT-MaSp, with all five predicted α -helices reasonably well superimposed (Figure 5B, 5C, and 5D).

While I-TASSER Model 3 was expected to yield a higher RMSD value due to its use of *ab initio* methods, its RMSD value remained relatively comparable to that of the template-based SWISS-MODEL and Phyre2 models (Table 3). However, the additional predicted α -helix between Trp3 and Ala19 in I-TASSER Model 3 did not superimpose with the corresponding structure in *E. australis* NT-MaSp, as the region exists as unmodeled loop in the experimentally determined template (Figure 5D). Both the structural quality and similarity assessments indicate that all three selected best models for predicting NT-AmSp structures, generated by SWISS-MODEL, Phyre2, and I-TASSER, adhered to parameter values set by ProSa-web, QMEAN, SAVES, and SuperPose, falling within established acceptable ranges. However, among the programs in SAVES, only the Verify3D scores did not apply to all three models due to the incompatibility of the NT-AmSp sequence with a relatively short length of fewer than 160 amino acids. These three structure prediction servers, employing distinct algorithms such as homology modeling, threading, and *ab initio* methods, have successfully produced *in silico* protein structures for NT-AmSp with relatively good quality. I-TASSER Model 3 was generated with the most complete amino acid sequence, exhibiting six α -helices with overall good quality values. The additional α -helical conformation represents the putative signal peptide region. In contrast, both SWISS-MODEL Model 1 and Phyre2 Model 1 display five α -helical structures across the N-terminus domain. The success of these predictions highlights the effectiveness of the employed modeling approaches in capturing the structural features of NT-AmSp.

Table 3

Superimposition of the predicted models with *E. australis* NT-MaSp1 from PDB

Server	Model	Superimposition of NT-AmSp1 with NT-MaSp1 (4FBS) (Å)
SWISS-MODEL	1	2.17
Phyre2	1	2.49
I-TASSER	3	2.04

Å = root mean square deviation (RMSD) value

Many previous studies on the tertiary structure of spidroin N-terminal region have elucidated the insights of the domain properties and function [11, 37-38]. Using the *in silico* approach, the tertiary structure of the monomeric N-terminal domain of the ampullate spidroin from *C. lyoni* prenymp was determined through computational tools including SWISS-MODEL, Phyre2, and I-TASSER, which are based on homology modeling, threading and *ab initio* methods. The structural prediction of NT-AmSp was performed on a monomeric basis rather than explicitly considering the homodimer state. Similar to other NT-domains of major ampullate spidroins, *C. lyoni* NT-AmSp was predicted to fold as an up-and-down (antiparallel) globular five-helix bundle (Figure 2A, 2C, 2E) that conforms to the stereotypical spidroin tertiary structure of five helical domains [11, 13, 39, 40-41]. As expected, all of the NT-AmSp predicted structures, threaded with high resolution, fit well with the solved MaSp1 from *E. australis* (4FBS), with RMSD scores ranging from 2.04 to 2.49. This result is consistent with the predicted structures of the end terminal domains of various spidroins, including the major ampullate, by homology-based prediction with RMSD scores between 0.38 and 2.5 [41]. The terminal

domains, including the N-terminus, are evolutionarily conserved, suggesting a highly similar function across all spidroins [13]. Upon synthesis in the tail of the major ampullate silk gland, the spidroin is stored as a highly concentrated semi-crystalline liquid protein known as dope in the sac at physiological pH [42]. At neutral pH, the N-terminal domain adopts a monomeric state with five α -helical conformations, conferring solubility to the spidroin. A decrease in pH as the liquid protein travels down to the spinning duct causes the protonation of the carboxylate side chains, leading to the dimerization of the N-terminus with a pKa of around 6.5. This stabilizes the NT and pulls the spidroin into a tight network [43]. To date, the N-terminal of spidroin has been observed to remain highly soluble in its helical form under various conditions [42].

In water-soluble proteins, approximately 35% of the total amino acid residues adopt the α -helical conformation [44]. This characteristic was consistently observed in all predicted structures of NT-AmSp in the present study. SWISS-MODEL generated Model 1 of NT-AmSp, featuring five helical bundles that cover 61.9% of the total amino acids. The alignment with the homologous template, MaSp1A (7wio.1.A) from *Triconephila clavipes* [45], revealed the highest sequence similarity of 42.4% compared to four other models predicted by the server (Table 1). Phyre2 also predicted the top model, Model 1, with five α -helices covering 65.8% of the total residues. This prediction was based on alignment with MaSp1 (c3lr6A) from *E. australis* [11], sharing 44.0% similarity. Meanwhile, I-TASSER utilized PSI-BLAST to identify related sequences, with similarity percentages ranging from 29% to 44%.

Subsequently, PSIPRED predicted secondary structures before threading the structures through the PDB structure library [23]. By integrating the ab initio method, Model 3 was generated, comprising a total of six α -helices covering 71.6% of the total residues. Notably, these included the same five helical bundles found in Model 1 by SWISS-MODEL and Phyre2. Surprisingly, the additional helical structure between Trp3 to Ala19, representing the putative signal peptide region, conformed to the typical structure of a signal peptide [46]. In many experimentally solved structures of MaSps using methods such as NMR and X-ray crystallography, the signal peptide region is often modelled as a random loop because the region is typically removed during protein translation in epithelial cells of the silk gland. It is worth noting that I-TASSER in this study stands out as a superior tool over homology and threading-based servers, providing extensive information on the NT-AmSp structure. Ab initio method is highly preferable when known template is unavailable [47].

Despite a high proportion of amino acid residues in the predicted α -helical conformations, not all of these helices are inherently hydrophilic, hydrophobic or amphipathic in nature. An α -helix represents the most common regular secondary structure in many water-soluble proteins [48]. Depending on the side chain chemical properties of amino acids, the α -helix can exhibit hydrophilicity or hydrophobicity with high number of polar residues or non-polar residues, respectively [49]. In some cases, it may display an amphipathic characteristic of being partially hydrophilic and hydrophobic [49]. However, to determine the nature of any of these helices in the predicted NT-AmSp, in-depth analysis is yet to be conducted. Nevertheless, the α -helices in the predicted NT-AmSp structures of all Model 1 and Model 3 were generated within similar amino acid regions, with only slight variations in the number of covered residues across the helix.

The predicted structure of the five α -helices of NT-AmSp are mainly hydrophobic. Helix-1 spans a range of 17 to 24 residues, with a high hydrophobic amino acid proportion of 66.7%, 58.3%, and 75.0% in Model 1 by SWISS-MODEL and Phyre2 and Model 3 by I-TASSER, respectively. Helix-2, on the other hand, was predicted in a 23-residual region with a high proportion of hydrophilic amino acids, constituting 65.2% in all models. Similar to Helix 1, Helix-3 possesses a considerable number of hydrophobic amino acids, ranging from 18 to 23 residues at 72.2%, 69.0%, and 71.4%, respectively.

Interestingly, both Helix-4 and Helix-5 exhibit a mixture of hydrophobic and hydrophilic amino acids in relatively similar ratios, ranging from 47.0% to 55.0% and 45.0% to 52.9%, respectively.

Both helical regions contain 17 to 23 amino acid residues in all models. Exclusively in Model 3, the α -helix core structure of the putative signal peptide region demonstrates an exceptionally long stretch of high hydrophobic amino acids, reaching up to 94.1%. It is a common feature found in all signal peptide motifs [50]. This result is corroborated by the previous hydropathy analysis by Kyte-Doolittle, showing that the N-terminal domain of AmSp is overall hydrophobic in nature, based on the calculated value of the grand average of hydropathicity (GRAVY) at 0.49 "(Mohtar), unpublished data)". GRAVY measures the hydrophobicity of protein in which positive and negative values indicate hydrophobicity and hydrophilicity. Similarly, the N-terminal domain of several spidroin types including tubuliform (TuSp) and pyriform (PySp) were also shown to display high amplitude of hydrophobicity [51,52]. Nevertheless, the analysis was unable to determine the percentage of residues buried in the hydrophobic core or hydrophilic residues exposed on the surface of the NT-AmSp (Figure 2B, 2D, 2F), which may contribute to understanding the orientation of the α -helices. The nature of the helices will remain vague unless determined experimentally.

From the present study, it can be inferred that the tertiary structure of the N-terminal ampullate spidroin from *C. lyoni* at an early developmental stage is highly conserved with that of orbicularian spiders in the adult stage. It is important to note that while these computational approaches provide valuable insights, the predictions are static and do not explicitly account for the protein being in a specific solution state, potentially limiting consideration of dynamic changes that could occur in a solution environment. *In silico* prediction tools with diversified algorithms, as used in this study, have proven beneficial as a preliminary window to provide insights into the three-dimensional structure of NT-AmSp from *C. lyoni* prenymp, which can be readily validated through advanced experimental methods such as Nuclear Magnetic Resonance (NMR) or X-ray Crystallography. With this information in hand, the predicted NT-AmSp structure can be employed to understand the functional diversity of the N-terminal domain of spidroin and utilized for the development of a soluble tag for *in vitro* or *in vivo* recombinant protein production [53].

4. Conclusions

The first top-ranked Model 1 by SWISS-MODEL and Phyre2, as well as Model 3 by I-TASSER with the highest C-score, were selected as the best models for the *in silico* prediction of the NT-AmSp structure. The quality of these models fulfils all the assessment parameters for a good model, including ProSA z-score, QMEAN4 score, percentage of ERRAT and total residues within the allowed region, overall G-factor, and RMSD values, except for Verify3D values, which were affected by the sequence's short length. Similar to other N-terminal spidroin domains, the NT-AmSp structure was predicted with five α -helices, with an additional helix representing the signal peptide region in Model 3. The *C. lyoni* NT-AmSp sequence exhibited an overall hydrophobic nature, and its structure was highly conserved with that of orbicularian spiders in the adult stage. The NT-AmSp structure represents the first N-terminal domain modelled from a non-orbicularian species at the prenympal stage.

Acknowledgement

The authors would like to thank the Faculty of Chemical Engineering & Technology, Universiti Malaysia Perlis (UniMAP) for the use of laboratory facilities. This work was supported by the Fundamental Research Grant Scheme (FRGS) under the grant number FRGS/1/2019/STG05/UNIMAP/03/1 from the Ministry of Education Malaysia.

References

- [1] Andersson, Marlene, Jan Johansson, and Anna Rising. "Silk spinning in silkworms and spiders." *International journal of molecular sciences* 17, no. 8 (2016): 1290. <https://doi.org/10.3390/ijms17081290>.
- [2] Brown, Cameron P., Alessandra D. Whaite, Jennifer M. MacLeod, Joanne Macdonald, and Federico Rosei. "With great structure comes great functionality: Understanding and emulating spider silk." *Journal of Materials Research* 30, no. 1 (2015): 108-120. <https://doi.org/10.1557/jmr.2014.365>.
- [3] Kerr, Genevieve G., Helen F. Nahrung, Aaron Wiegand, Joanna Kristoffersen, Peter Killen, Cameron Brown, and Joanne Macdonald. "Mechanical properties of silk of the Australian golden orb weavers *Nephila pilipes* and *Nephila plumipes*." *Biology Open* 7, no. 2 (2018): bio029249. <https://doi.org/10.1242/bio.029249>.
- [4] Gosline, John M., Paul A. Guerette, Christine S. Ortlepp, and Ken N. Savage. "The mechanical design of spider silks: from fibroin sequence to mechanical function." *Journal of Experimental Biology* 202, no. 23 (1999): 3295-3303. <https://doi.org/10.1242/jeb.202.23.3295>.
- [5] Sponner, Alexander, Wolfram Vater, Shamci Monajembashi, Eberhard Unger, Frank Grosse, and Klaus Weisshart. "Composition and hierarchical organisation of a spider silk." *PloS one* 2, no. 10 (2007): e998. <https://doi.org/10.1371/journal.pone.0000998>.
- [6] Ramezaniaghdam, Maryam, Nadia D. Nahdi, and Ralf Reski. "Recombinant spider silk: promises and bottlenecks." *Frontiers in Bioengineering and Biotechnology* 10 (2022): 835637. <https://doi.org/10.3389/fbioe.2022.835637>.
- [7] Garb, Jessica E., Robert A. Haney, Evelyn E. Schwager, Matjaž Gregorič, Matjaž Kuntner, Ingi Agnarsson, and Todd A. Blackledge. "The transcriptome of Darwin's bark spider silk glands predicts proteins contributing to dragline silk toughness." *Communications biology* 2, no. 1 (2019): 275. <https://doi.org/10.1038/s42003-019-0496-1>.
- [8] Blamires, Sean J., Todd A. Blackledge, and I-Min Tso. "Physicochemical property variation in spider silk: ecology, evolution, and synthetic production." *Annual review of entomology* 62, no. 1 (2017): 443-460. <https://doi.org/10.1146/annurev-ento-031616-035615>.
- [9] Matsuhira, Takashi, and Shigeyoshi Osaki. "Molecular weight of *Nephila clavata* spider silk." *Polymer journal* 47, no. 6 (2015): 456-459. <https://doi.org/10.1038/pj.2015.10>.
- [10] Chen, Xin, Zhengzhong Shao, and Fritz Vollrath. "The spinning processes for spider silk." *Soft Matter* 2, no. 6 (2006): 448-451. <https://doi.org/10.1039/b601286h>.
- [11] Askarieh, Glareh, My Hedhammar, Kerstin Nordling, Alejandra Saenz, Cristina Casals, Anna Rising, Jan Johansson, and Stefan D. Knight. "Self-assembly of spider silk proteins is controlled by a pH-sensitive relay." *Nature* 465, no. 7295 (2010): 236-238. <https://doi.org/10.1038/nature08962>.
- [12] Šede, Megija, Jēkabs Fridmanis, Martins Otkovs, Jan Johansson, Anna Rising, Nina Kronqvist, and Kristaps Jaudzems. "Solution structure of tubuliform spidroin N-terminal domain and implications for pH dependent dimerization." *Frontiers in Molecular Biosciences* 9 (2022): 936887. <https://doi.org/10.3389/fmolb.2022.936887>.
- [13] Rising, Anna, Göran Hjälrm, Wilhelm Engström, and Jan Johansson. "N-terminal nonrepetitive domain common to dragline, flagelliform, and cylindriciform spider silk proteins." *Biomacromolecules* 7, no. 11 (2006): 3120-3124. <https://doi.org/10.1021/bm060693x>.
- [14] Heiby, Julia C., Benedikt Goretzki, Christopher M. Johnson, Ute A. Hellmich, and Hannes Neuweiler. "Methionine in a protein hydrophobic core drives tight interactions required for assembly of spider silk." *Nature communications* 10, no. 1 (2019): 4378. <https://doi.org/10.1038/s41467-019-12365-5>.
- [15] Chaw, Ro Crystal, Christopher A. Sasaki, and Cheryl Y. Hayashi. "Complete gene sequence of spider attachment silk protein (PySp1) reveals novel linker regions and extreme repeat homogenization." *Insect Biochemistry and Molecular Biology* 81 (2017): 80-90. <https://doi.org/10.1016/j.ibmb.2017.01.002>.
- [16] Yarger, Jeffery L., Brian R. Cherry, and Arjan Van der Vaart. "Uncovering the structure–function relationship in spider silk." *Nature Reviews Materials* 3, no. 3 (2018): 1-11. <https://doi.org/10.1038/natrevmats2018.8>.
- [17] Correa-Garhwal, Sandra M., and Jessica E. Garb. "Diverse formulas for spider dragline fibers demonstrated by molecular and mechanical characterization of spitting spider silk." *Biomacromolecules* 15, no. 12 (2014): 4598-4605. <https://doi.org/10.1021/bm501409n>.
- [18] Gatesy, John, Cheryl Hayashi, Dagmara Motriuk, Justin Woods, and Randolph Lewis. "Extreme diversity, conservation, and convergence of spider silk fibroin sequences." *Science* 291, no. 5513 (2001): 2603-2605. <https://doi.org/10.1126/science.1057561>.
- [19] Arakawa, Kazuharu, Nobuaki Kono, Ali D. Malay, Ayaka Tateishi, Nao Ifuku, Hiroyasu Masunaga, Ryota Sato et al. "1000 spider silkomes: Linking sequences to silk physical properties." *Science advances* 8, no. 41 (2022): eabo6043. <https://doi.org/10.1126/sciadv.abo6043>.

- [20] Waterhouse, Andrew, Martino Bertoni, Stefan Bienert, Gabriel Studer, Gerardo Tauriello, Rafal Gumieny, Florian T. Heer et al. "SWISS-MODEL: homology modelling of protein structures and complexes." *Nucleic acids research* 46, no. W1 (2018): W296-W303. <https://doi.org/10.1093/nar/gky427>.
- [21] Kelley, Lawrence A., Stefans Mezulis, Christopher M. Yates, Mark N. Wass, and Michael JE Sternberg. "The Phyre2 web portal for protein modeling, prediction and analysis." *Nature protocols* 10, no. 6 (2015): 845-858. <https://doi.org/10.1038/nprot.2015.053>.
- [22] Yang, Jianyi, Renxiang Yan, Ambrish Roy, Dong Xu, Jonathan Poisson, and Yang Zhang. "The I-TASSER Suite: protein structure and function prediction." *Nature methods* 12, no. 1 (2015): 7-8. <https://doi.org/10.1038/nmeth.3213>.
- [23] Zhou, Xiaogen, Wei Zheng, Yang Li, Robin Pearce, Chengxin Zhang, Eric W. Bell, Guijun Zhang, and Yang Zhang. "I-TASSER-MTD: a deep-learning-based platform for multi-domain protein structure and function prediction." *Nature protocols* 17, no. 10 (2022): 2326-2353. <https://doi.org/10.1038/s41596-022-00728-0>.
- [24] Wiederstein, Markus, and Manfred J. Sippl. "ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins." *Nucleic acids research* 35, no. suppl_2 (2007): W407-W410. <https://doi.org/10.1093/nar/gkm290>.
- [25] Benkert, Pascal, Marco Biasini, and Torsten Schwede. "Toward the estimation of the absolute quality of individual protein structure models." *Bioinformatics* 27, no. 3 (2011): 343-350. <https://doi.org/10.1093/bioinformatics/btq662>.
- [26] Dym Ora, Eisenberg David and Yeates Todd O. (2012). Detection of errors in protein models. In *International Tables for Crystallography Volume F: Crystallography of Biological Macromolecules*, edited by E. Arnold, D. M. Himmel, and M. G. Rossmann, 2nd ed., Indianapolis: John Wiley, Page 677–679.
- [27] Maiti, Rajarshi, Gary H. Van Domselaar, Haiyan Zhang, and David S. Wishart. "SuperPose: a simple server for sophisticated structural superposition." *Nucleic acids research* 32, no. suppl_2 (2004): W590-W594. <https://doi.org/10.1093/nar/gkh477>.
- [28] Jaudzems, Kristaps, Glareh Askarieh, Michael Landreh, Kerstin Nordling, My Hedhammar, Hans Jörnvall, Anna Rising, Stefan D. Knight, and Jan Johansson. "pH-Dependent dimerization of spider silk N-Terminal domain requires relocation of a wedged tryptophan side chain." *Journal of molecular biology* 422, no. 4 (2012): 477-487. <https://doi.org/10.1016/j.jmb.2012.06.004>.
- [29] Yadav, Brijesh S., Vijay Tripathi, Ajeet Kumar, Md Faheem Khan, Abhijit Barate, Ajay Kumar, and Bhaskar Sharma. "Molecular modeling and docking characterization of Dectin-1 (PAMP) receptor of Bubalus bubalis." *Experimental and molecular pathology* 92, no. 1 (2012): 7-12. <https://doi.org/10.1016/j.yexmp.2011.09.018>.
- [30] Gupta, Chhedi Lal, Salman Akhtar, Preeti Bajpaib, K. N. Kandpal, G. S. Desai, and Ashok K. Tiwari. "Computational modeling and validation studies of 3-D structure of neuraminidase protein of H1N1 influenza A virus and subsequent in silico elucidation of piceid analogues as its potent inhibitors." *EXCLI journal* 12 (2013): 215.
- [31] Benkert, Pascal, Silvio CE Tosatto, and Dietmar Schomburg. "QMEAN: A comprehensive scoring function for model quality assessment." *Proteins: Structure, Function, and Bioinformatics* 71, no. 1 (2008): 261-277. <https://doi.org/10.1002/prot.21715>.
- [32] Messaoudi, Abdelmonaem, Hatem Belguith, and Jeannette Ben Hamida. "Homology modeling and virtual screening approaches to identify potent inhibitors of VEB-1 β -lactamase." *Theoretical Biology and Medical Modelling* 10, no. 1 (2013): 22. <https://doi.org/10.1186/1742-4682-10-22>.
- [33] Singh, Ravinder, Ankita Gurao, C. Rajesh, S. K. Mishra, Saroj Rani, Ankita Behl, Vikash Kumar, and R. S. Kataria. "Comparative modeling and mutual docking of structurally uncharacterized heat shock protein 70 and heat shock factor-1 proteins in water buffalo." *Veterinary world* 12, no. 12 (2019): 2036. <https://doi.org/10.14202/vetworld.2019.2036-2045>.
- [34] Tran, Ngoc Tuan, Ivan Jakovlić, and Wei-Min Wang. "In silico characterisation, homology modelling and structure-based functional annotation of blunt snout bream (*Megalobrama amblycephala*) Hsp70 and Hsc70 proteins." *Journal of Animal Science and Technology* 57, no. 1 (2015): 44. <https://doi.org/10.1186/s40781-015-0077-x>.
- [35] Lovell Simon C., Davis Ian W., Arendall W. Bryan III, de Bakker Paul I. W., Word Jimin M., Prisant Michael G., Richardson Jane S. and Richardson David C. "Structure validation by Calpha geometry: Phi, psi and C-beta deviation." *Proteins* 50, no. 3 (2003): 437–450. <https://doi.org/10.1002/prot.10286>.
- [36] Ramírez, David, and Julio Caballero. "Is it reliable to take the molecular docking top scoring position as the best solution without considering available structural data?." *Molecules* 23, no. 5 (2018): 1038. <https://doi.org/10.3390/molecules23051038>.
- [37] Landreh, Michael, Glareh Askarieh, Kerstin Nordling, My Hedhammar, Anna Rising, Cristina Casals, Juan Astorga-Wells et al. "A pH-dependent dimer lock in spider silk protein." *Journal of molecular biology* 404, no. 2 (2010): 328-336. <https://doi.org/10.1016/j.jmb.2010.09.054>.

- [38] Bauer, Joschka, and Thomas Scheibel. "Conformational stability and interplay of helical N-and C-terminal domains with implications on major ampullate spidroin assembly." *Biomacromolecules* 18, no. 3 (2017): 835-845. <https://doi.org/10.1021/acs.biomac.6b01713>.
- [39] Bauer, Joschka, Daniel Schaal, Lukas Eisoldt, Kristian Schweimer, Stephan Schwarzing, and Thomas Scheibel. "Acidic residues control the dimerization of the N-terminal domain of black widow spiders' major ampullate spidroin 1." *Scientific reports* 6, no. 1 (2016): 34442. <https://doi.org/10.1038/srep34442>.
- [40] Otikovs, Martins, Gefei Chen, Kerstin Nordling, Michael Landreh, Qing Meng, Hans Jörnval, Nina Kronqvist, Anna Rising, Jan Johansson, and Kristaps Jaudzems. "Diversified structural basis of a conserved molecular mechanism for pH-dependent dimerization in spider silk N-terminal domains." *ChemBioChem* 16, no. 12 (2015): 1720-1724. <https://doi.org/10.1002/cbic.201500263>.
- [41] Collin, Matthew A., Thomas H. Clarke III, Nadia A. Ayoub, and Cheryl Y. Hayashi. "Genomic perspectives of spider silk genes through target capture sequencing: Conservation of stabilization mechanisms and homology-based structural models of spidroin terminal regions." *International Journal of Biological Macromolecules* 113 (2018): 829-840. <https://doi.org/10.1016/j.ijbiomac.2018.02.032>.
- [42] Andersson, Marlene, Gefei Chen, Martins Otikovs, Michael Landreh, Kerstin Nordling, Nina Kronqvist, Per Westermarck et al. "Carbonic anhydrase generates CO₂ and H⁺ that drive spider silk formation via opposite effects on the terminal domains." *PLoS biology* 12, no. 8 (2014): e1001921. <https://doi.org/10.1371/journal.pbio.1001921>.
- [43] Kronqvist, Nina, Martins Otikovs, Volodymyr Chmyrov, Gefei Chen, Marlene Andersson, Kerstin Nordling, Michael Landreh et al. "Sequential pH-driven dimerization and stabilization of the N-terminal domain enables rapid spider silk formation." *Nature communications* 5, no. 1 (2014): 3254. <https://doi.org/10.1038/ncomms4254>.
- [44] Martin, Juliette, Guillaume Letellier, Antoine Marin, Jean-François Taly, Alexandre G. De Brevern, and Jean-François Gibrat. "Protein secondary structure assignment revisited: a detailed analysis of different assignment methods." *BMC structural biology* 5, no. 1 (2005): 17. <https://doi.org/10.1186/1472-6807-5-17>.
- [45] Oktaviani, Nur Alia, Ali D. Malay, Akimasa Matsugami, Fumiaki Hayashi, and Keiji Numata. "Unusual pK_a Values Mediate the Self-Assembly of Spider Dragline Silk Proteins." *Biomacromolecules* 24, no. 4 (2023): 1604-1616. <https://doi.org/10.1021/acs.biomac.2c01344>.
- [46] Garnier, Jean, Pierre Gaye, Jean-Claude Mercier, and Barry Robson. "Structural properties of signal peptides and their membrane insertion." *Biochimie* 62, no. 4 (1980): 231-239. [https://doi.org/10.1016/s0300-9084\(80\)80397-x](https://doi.org/10.1016/s0300-9084(80)80397-x).
- [47] Haron, F., A. Azazi, K. Chua, Y. Lim, P. Lee, and C. Chew. "RESEARCH ARTICLE In silico structural modeling and quality assessment of Plasmodium knowlesi apical membrane antigen 1 using comparative protein models." *Trop. Biomed* 39 (2022): 394-401. <https://doi.org/10.47665/tb.39.3.009>.
- [48] Zhang, Shao-Qing, Daniel W. Kulp, Chaim A. Schramm, Marco Mravic, Ilan Samish, and William F. DeGrado. "The membrane-and soluble-protein helix-helix interactome: similar geometry via different interactions." *Structure* 23, no. 3 (2015): 527-541. <https://doi.org/10.1016/j.str.2015.01.009>.
- [49] Qing, Rui, Shilei Hao, Eva Smorodina, David Jin, Arthur Zalevsky, and Shuguang Zhang. "Protein design: From the aspect of water solubility and stability." *Chemical Reviews* 122, no. 18 (2022): 14085-14179. <https://doi.org/10.1021/acs.chemrev.1c00757>.
- [50] Yuan, Zheng, Melissa J. Davis, Fasheng Zhang, and Rohan D. Teasdale. "Computational differentiation of N-terminal signal peptides and transmembrane helices." *Biochemical and biophysical research communications* 312, no. 4 (2003): 1278-1283. <https://doi.org/10.1016/j.bbrc.2003.11.069>.
- [51] Wang, Kangkang, Rui Wen, Qiupin Jia, Xiangqin Liu, Junhua Xiao, and Qing Meng. "Analysis of the full-length pyriform spidroin gene sequence." *Genes* 10, no. 6 (2019): 425. <https://doi.org/10.3390/genes10060425>.
- [52] Wen, Rui, Xiangqin Liu, and Qing Meng. "Characterization of full-length tubuliform spidroin gene from *Araneus ventricosus*." *International Journal of Biological Macromolecules* 105 (2017): 702-710. <https://doi.org/10.1016/j.ijbiomac.2017.07.086>.
- [53] Kronqvist, Nina, Médoune Sarr, Anton Lindqvist, Kerstin Nordling, Martins Otikovs, Luca Venturi, Barbara Pioselli et al. "Efficient protein production inspired by how spiders make silk." *Nature communications* 8, no. 1 (2017): 15504. <https://doi.org/10.1038/ncomms15504>.