Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Structural Elucidation of the Masp1 and Masp2 Protein from Nephila Pilipes Web Via Bioinformatics Approaches

Anwardi Jamil[1], Amirul Ridzuan Abu Bakar[1*], Nurul Husna Khairuddin[2]

[1] Faculty of Chemical Engineering Technology, Universiti Malaysia Perlis, Jejawi, 02600, Perlis, Malaysia
[2] Faculty of Medicine and Health Sciences, University Tunku Abdul Rahman, Bandar Sungai Long, Kajang, 43000, Selangor, Malaysia

ARTICLE INFO

Article history:
Received 7 October 2022
Received in revised form 19 December 2022
Accepted 17 February 2023
Available online 11 March 2023

Keywords:
Nephila Pilipes; In Silico; 3D Modeling; Bioinformatics

ABSTRACT

Major ampullate spidroin (MaSps) from orb-weaver spp spider has recently gained interest due to its exceptional characteristics. The biomechanical and biochemical properties from MaSps offer potential in harvesting and exploiting MaSps as a promising bio-based product. However, the current research on the structural elucidation focused more onto the *Nephila clavipes* spider web rather than the Nephila pilipes which are more common in this region. Herein, this study integrates the used of computational power and algorithm to elucidate the 3D protein morphology of MaSp1 and MaSp2 of Nephila pilipes dragline silk protein using nearly complete amino acid sequences obtained from the protein database (PDB). In silico homology modelling via Phyre2, SWISS-MODEL and RaptorX was adopted to predict the protein structure of MaSP-1 and 2 using proteins threading, automated comparative modelling of three-dimensional (3D) protein structures and deep learning approaches. Consequently, we described a thorough 3D protein models of MaSp1 and MaSp2 with a higher percentage of coils, α-helix and a low percentage of β-sheet on repetitive regions of MaSp1 and MaSp2. The results of this current work provide insights into Bioinformatics potentials in engineering spider silk-based biomaterial and bridging the most apparent gaps in the knowledge of MaSp1 and MaSp2.

1. Introduction

Spider silk protein or spidroin is known for its impressive mechanical properties, having one of the strongest biomaterials that outmatched the mechanical properties possessed by both steel and Kevlar as mentioned by Whittall *et al.,* [1]. Interestingly, the assembly of spider silk protein contains one of the most valuable biological polymers and remarkable biomimetic potentials including the sturdiness of web constructions, containing abundant amino acid and carboxyl groups, and specific purpose design that could contribute to many biotechnological applications as suggested by Kiew *et al.,* [2]. In brief, the manufacturing of silk proteins was formed inside the major ampullate gland of a spider whereby the major ampullate spidroin (MaSp) one and two were used in constructing the web frames, lifeline, and used to capture prey as studied by Hajer and Řeháková [3]. The specific

* *Corresponding author.*
E-mail address: amirulridzuan@unimap.edu.my

https://doi.org/10.37934/araset.30.1.132144

characteristics of spidroins have become the most common research to unravel the molecular structure of spidroin, which contains a great variety of mechanical, physicochemical, and biological properties as reviewed by Fai *et al.,* [4].

Currently, the implementation of technologies has become a breakthrough in understanding the molecular structure of the protein. However, the process is still progressively slow due to the abundance of new evolutionary species. In addition, as the research on spider webs evolved due to the acclimatization and surrounding needs, the complete amino acid sequence of spider silk protein is still insufficient. This was proven by a finding from José Roberto that stated the molecular sequence, as well as the structure of the spidroin, was partially documented and characterized by José Roberto *et al.*, [5]. Apart from this, the limitations arose when limited in-depth bioprospecting studies were made to understand the spidroin of local Nephilid species even though Malaysia was blessed with high biodiversity of this species thus, hindering the possibilities of innovation to emerge.

In general, the *Nephila pilipes* (also known as the northern golden orb weaver) is a species of golden orb-web spider that is mostly inhabited in East and Southeast Asia countries. These species of *N. pilipes* were habitually found in primary and secondary forests and gardens. The silk fiber from *N. pilipes* has a unique structure spidroin that can be brought up as a possible biomaterial as mentioned by Abdullah *et al.,* [6]. This indicates promising opportunities for the development of novel biofibers and the discovery of more diversified proteins compose of spidroin. Previous studies show *N. pilipes* amino acid arrangements that are different from the aforementioned *Nephilia clavipes* with only minor similarities. Further understanding of the 3D structure models for spidroin is impossible without in-depth knowledge of the protein's secondary structure, thus, preventing the potential use of the fiber from a wider perspective. Consequently, the retrieval of the silk protein structure is initially crucial to provide concrete data for future protein design as described by Rim *et al.*, [7]. Thorough investigations should be conducted to bridge the existing gap. Thus, the present study seeks to divulge the predicted three-dimensional structure protein modeling of the MaSp1 and MaSp2 from *N. pilipes* web for future biotechnological applications.

A study by Abu Bakar *et al.*, [8] have shown that by incorporating Bioinformatics to understand the capabilities in construing molecular behavior of spider silk protein, the objective in this research was focused on the prediction and modeling of the 3D structure of proteins using Bioinformatics to thoroughly elucidate the protein secondary structure of MaSp1 and MaSp2 of *N. pilipes*. Through the execution of multiple computational algorithms, this technique utilized protein threading, comparative and deep learning approaches to elucidate and predicts the 3D structure of a query protein through the sequence alignment of template proteins. A tertiary structure from a full sequence of amino acids containing N-terminal, C-terminal and the repetitive domain obtained from the protein database (PDB) *N. pilipes* was identified and modelled to acquire the overall view of spidroin structure. As a final result, the 3D protein structure of spidroin was modelled (in silico) using Phyre2, SWISS-MODEL and RaptorX through a comparative modelling approach and viewed using UCSF Chimera.

2. Methodology

2.1 In Silico Structural Elucidation of MaSp1 and MaSp2 Proteins

The sequencing data were obtained through the Protein database (PDB) on *N. pilipes* MaSp1 and MaSp2 amino acid. It was then elucidated using Protein Homology/analogY Recognition Engine ver. 2 (Phyre2), SWISS-MODEL and RaptorX software to obtain the 3D structure model. The amino acid sequence of MaSp1 and MaSp2 were converted as FASTA format (text-based format) and later computes to Phyre2, Swissmodel, and RaptorX-ContactMap programs. The analysis template

database and the predicted secondary structure produced from Phyre2, SWISS-MODEL and RaptorX were further interpreted before constructing the 3D protein structure models. For visualization and overlaying of the 3D protein structures and amino acid localization software, namely UCSF Chimera, was used.

2.1.1 Phyre2 analysis

The complete amino acid sequence of MaSp1 and MaSp2 in FASTA format was introduced to the Phyre2 algorithm to construct a 3D protein structure modelling. These peptide sequences of MaSp1 and MaSp2 were input in the Phyre2 via http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index. Once the peptide sequence was entered in the amino acid sequence column, users had the option to select either "Normal" or "Intensive" modelling mode for protein modelling for the interface. By default, in this experiment, the modelling mode was set to "Normal" and "Other". The predicted model was allocated as WinRAR files containing all the necessary information needed to analyses the 3D structure.

2.1.2 SWISS-MODEL analysis

The MaSp1 and MaSp2 sequences in FASTA format were introduced to the SWISS-MODEL server (https://swissmodel.expasy.org) to generate a predicted 3D protein structure model. On this server, the predicted models were constructed using algorithm and selection comparison. When the 3D modelling was completed, the results were made accessible in WinRAR files, containing all the data required was analyzed and inspected further into each predicted structure.

2.1.3 RaptorX analysis

The selected data obtained from the sequencing as FASTA formats were introduced to the RaptorX server (http://raptorx.uchicago.edu). Predictive models were built based on algorithm and selection comparison on this server. The outcomes as WinRAR files were accessible when the built-up of the 3D modelling was completed, providing all the data required was analyzed and look deeper for each predicted structure.

3. Results and Discussions

3.1 In Silico Peptide Sequence Analysis

The purpose of in silico modelling was to identify the 3D protein structure of MaSp1 and MaSp2 through homology or template-based modelling technique using different approaches such as using Phyre2, SWISS-MODEL and RaptorX. It was based on three categories; comparative modelling either by homology or template-based modelling, protein-protein docking and hybrid/integrative structure modelling in review paper by several authors [9-10]. Template-based protein structure modelling techniques focus on the study of rules that dictate the 3D structure of natural proteins from the point of view of evolution theory. To focus on template-based structure modelling, methods for template-based structure modelling was analyzed by evaluating techniques available for all the main steps involved in the comparative modelling process. Integrative structure modelling (ISM) aims to obtain structural insights of protein structural models using a computational method by combining various information from multiple techniques to achieve better accuracy, resolution and precision as mentioned by Braitbard *et al.,* [11]. Before the computational approach can be done, suitable

templates must first be discovered through simple pairwise sequence alignment methods such as SSEARCH (local), Basic Local Alignment Search Tool (BLAST) and FASTA as recommended by Webb and Sali [12].

Initially, the amino acid from Protein database (PDB) was converted into FASTA format for the computational modelling starting point. The FASTA format is referred as text-based representing either a nucleotide sequence or peptide sequence that indicates the base pairs or amino acids (protein) using a single latter code. The format allows for the sequence names and comments to precede the sequences. In Bioinformatics, the FASTA format originated from a FASTA computational algorithm package that becomes a universal standard to identify 3D structure protein of interest. The advantages of the FASTA format were its' simplicity and ease to operate using text-processing tools and scripting languages like R programming language, Python, Ruby, and Perl as reviewed by several authors [13-15]. A sequence of FASTA format begins with a single-line description followed by lines of sequence data. The collections of MaSps protein were obtained from PDB database and were later converted as FASTA files format. The sequence of the protein was depicted as Figure 1 and Figure 2.

```
>seq
AGQGGAGAAAGGAGQGGYGGLGGQGAGQGGSSKLQALNMANSFISAASNSAGGAGQGGYG
GLGSQGAGRGGQGAGAAAAADQTFLNEIASGLVKMFSQRGGLGGQGAGAAAAAAAGGAGQ
GGYGGLGNQGAGRGGAGAAATSIGGTGQGGYGGLGSQGAGRGGLGGQGAGAAAAAAGGAG
QGGYGGAGGYTAGAGGYGGLGSQAGVRGGLGGQGAGAAAAAAAGGAGQGGLGGQGAGQGA
GASAAAAGAAGQGGYGGLGSQGAGSYGEGAGAAAAAAGGAGQGGYGGLGGQGAGQGGYGG
LGSQGAGRGGLGGQGYGAAAPGQGAGATAGGQGAGQGAGAGLGGQGGAGQRGAAGLGSQG
AGRGGLGGQGAGAVAAAAAGGAGQGGYGGLGAGQGGYGRGGAGQAAAAAGGAGQRGYGGL
GNQGAGYGGLGADSAGAAAGAAYGGGGQGGYGGRGPGGYGAATGAGPGGYGGAGPGAGAG
LGSQGAGRGGQGAGAAAAAAAPAGQEGIRGQGAGQGGYGGLGSQGSGAGGLGGQGAGAAA
AAAGGAGQGGLGGQGAGQGAGAAAAAAGGVRQGGYGGLGSQGAGRGGQGAGSAGGTAQGP
GQGGYGGLGGQGVGRGGLGGQGAGAAAAGGAGQGGYGGVGSGASAASAAASLRSSPQASS
AGQGGQGGYGAYGRGAGAALAATGVGPVGQIGASNPGLSGCDVLIQALLEVVSALIQILG
SSSIGQVNYGSAGQATQIVGQSVYQALG
```

Fig. 1. FASTA format for Major Ampullate Spidroin 1 (MaSp1)

```
>protein
GGPGQQGPGGPGGQGPGRQGPSGPGSAAAAAGAGQQGPGGAFIGAFMNAAYGPSQSGAFR
YGPGQQGPSGSNSKAAAAAGTQSKLQALNMDMAFASGGYGQGQQGPSMDGVKTNAGPGQQ
AESGQQGPGGYGPGQVNEMRSLPGQQGPAAVNEVSYGGGAPGSAAAAAAAASGPSGYGPG
AAPGGSGPGGGYGQGPQGPSGGGSAAAAAAAASGPGQQGPSGYGPGQQGPGAAAGRQQGG
SGPGSAAAAAAAGPGGYGPGGPRQQGPGGYGPGGSGAAAAAAAGPGGYGPGQQGPGGYGP
AGAAPAAAAAGSGPGGYGPGPRGPGAAAAGPGGYGPGQQGASAAASAAAGGAGGAGGYAG
PGQQGPGGYGPGQQGPSGPGSASAAAAAAAAGPGGYGPGQGPGGAGAAATAAAAAGPGGY
AAAAAAAAGPGAAAAGAAAAGGSGAAAAAAAAAGGAGPGRAAAAGPGGYGPAQQGPSGPG
IAASAASYGPGGAGPAQQGPAGYGPGSAVAAGPGGSGAAAGPGSQASAAASRLASPDSGA
RVASAVSNLVSSGPTSSAALSSVISNAVSQIGASNPGLSGCDVLIQALLEIVSACVTILS
SSSIGQVNYGPGGAGAAAAAAAGGAGPGRQQEYG
```

Fig. 2. FASTA format for Major Ampullate Spidroin 2 (MaSp2)

The sequence was submitted in FASTA format for all 3D structure modelling structure computational algorithm (Phyre2, SWISS-MODEL and RaptorX) to identify the molecular structure of MaSp1 and MaSp2. The advantage of using this format is more sensitive derivation of the FASTP can be achieved, which can search protein or DNA sequence from the database, compare protein sequence to a DNA sequence database and translating the protein and DNA sequence. FASTA provides unique capabilities compared to others as it can be used to search sequence databases, evaluate similarity scores, and identify periodic structures depending on the local sequence similarity. Additionally, FASTA computes the calculation of the initial pairwise similarity score that compares the protein sequences, allowing multiple regions of similarities were joined together increasing the score matrices based on the related sequence from the database. The sensitivity and selectivity in biological sequence comparison depend on the selection of sequence; the more related sequences will increase the sensitivity contrary to the unrelated sequences. Thus, the improved algorithm (FASTA algorithm) through the heuristic approach was used to increase the speed of the process while obtaining as much sensitivity as possible the amino acid as reviewed by several authors [16-19]. FASTA format is famous for its capability as a preferred approach paired with other sequence processing tools such as Basic Local Alignment Search Tool (BLAST). BLAST is the local alignment algorithm used templates database through the pairwise sequence–sequence comparisons to pin-point the region of similarity between biological sequences as mentioned by Fiser [19]. With the combination between BLAST and global alignment algorithm, the full primer-target alignment can be accomplished and sensitive to detect targets that with a significant number of mismatches to primers.

3.2 Phyre2 Structural Predictions of MaSp Proteins

Phyre2 is a protein structure prediction server that displays a molecular structure provided with an easy and simple intuitive interface with state-of-the-art protein Bioinformatics tools as suggested by Kelley *et al.,* [20]. The function of Phyre2 predicts the structure of a query protein using an amino acid sequence. The model prediction was reconstructed by using recognize template from a homologous structure. The reported results that were analyzed by the Phyre2 algorithms were composed of top model info, secondary structure/disorder, domain analysis and detailed template information. The outcomes of Phyre2 indicate the top-scoring model (confidence score and coverage) and the dimensions of the protein was shown as Angstrom unit (Å). Furthermore, the predicted top model can also be viewed using 'interactive 3D view in JSmol' that acts as a 3D molecular visualization tool. The structural elucidation and figures obtained from Phyre2 protein prediction algorithm for MaSps 1 and 2 were analyzed thoroughly and illustrated as Figure 3.

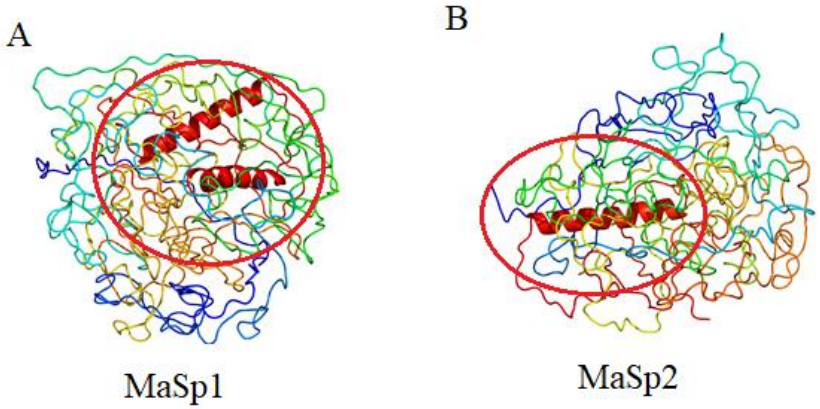

Fig. 3. Final model prediction of Major Ampullate Spidroin 1 (MaSp1) and 2 (MaSp2) by Phyre2

The predicted secondary structure and disorder were based on the amino acid residue that represented by colour coded (yellow, green, red and purple) according to the simple properties scheme: (A, S, T, G, P: small/polar) was represented in yellow, (M, I, L, V: hydrophobic) was represented in green, (K, R, E, N, D, H, Q: charged) was represented in red and (W, Y, F, C: aromatic + cysteine) was represented in purple in which the sequence was then interpreted into α-helix, β-sheet and coil regions. The domain analysis act as an indicator for determining the protein region that has been modelled based on the amino acid sequence with colored rows in which the colours represent the confidence of the homology (i.e., high confidence represent red and low confidence is represented in blue). The detailed template information summarizes the displays information based on template code, alignment coverage, 3D model, confidence, percentage sequence identity and the templates protein information as described by Kelley *et al.,* [20]. The predicted model for MaSp1 shows 7 % of residues model from the amino acids residues and has several regions with high and moderate confidence keys, respectively. Whereas, the predicted model for MaSp2 showed less than 1% of residues model from the amino acids residues and had some regions with moderate confidence key. Nonetheless, Figure 3 (a) and Figure 3 (b) had a limitation due to the high percentage of predicted disorder regions that disrupts the whole protein structure prediction.

3.3 SWISS-MODEL Structural Predictions of MaSp Proteins

SWISS-MODEL utilized a fully automated protein structure homology-modelling algorithms that provide accessibility for protein modelling including the modelling of homo- and heteromeric complexes as revised by Waterhouse *et al.,* [21]. The identified amino acid sequences in FASTA format initiate the starting point for protein modelling analysis and template base searching using the SWISS-MODEL template library (SMTL). The function of the SMTL was to detect the related evolutionary structure matching from the target sequence (MaSp1 and MaSp2) using the BLAST and Bioinformatics tools kit (HHBlits) as highlighted by several authors [22-24]. The structural elucidation and figures obtained from SWISS-MODEL protein prediction algorithm for MaSps 1 and 2 were analyzed thoroughly and illustrated as Figure 4 and Figure 5.

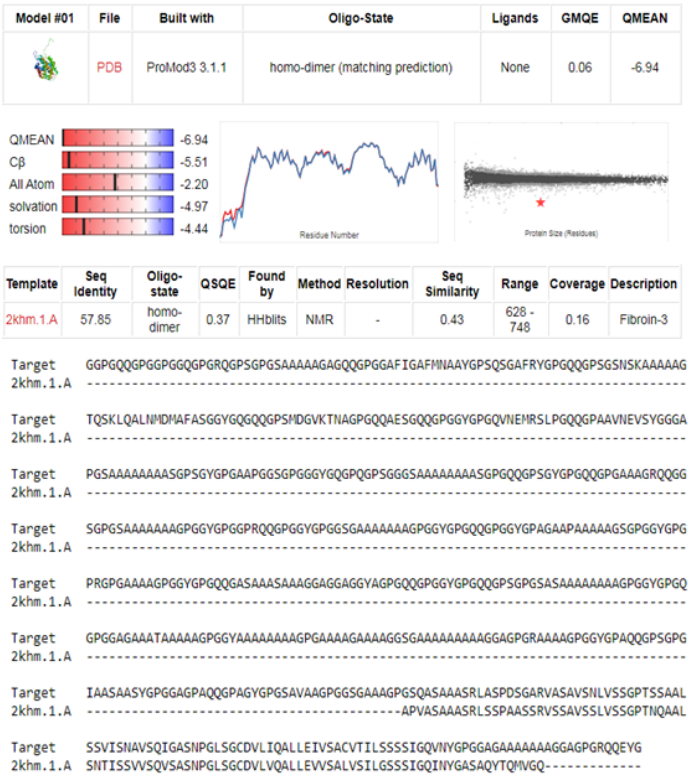

Fig. 4. Predicted model and template for MaSp1 using SWISS-MODEL

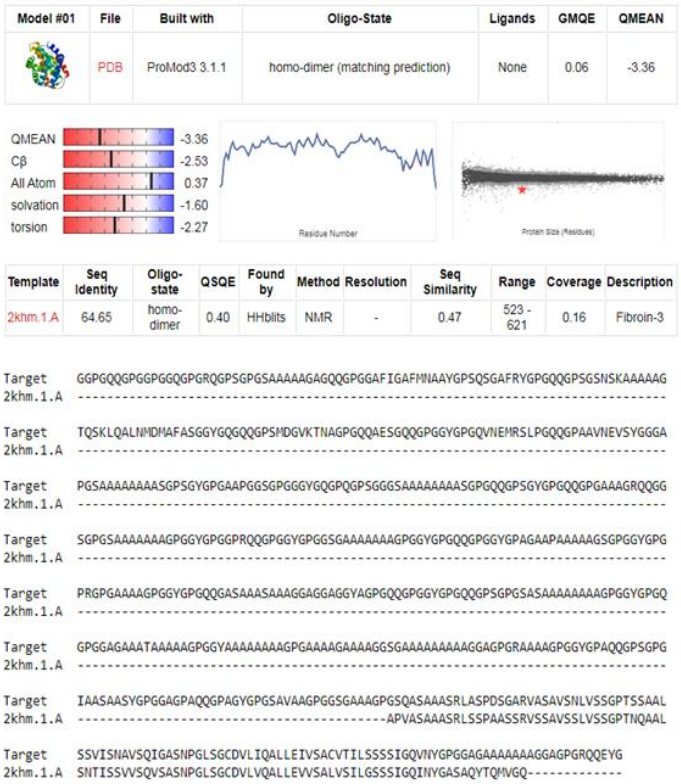

Fig. 5. Predicted model and template for MaSp2 using SWISS-MODEL

The overall template found and reported by SMTL showed a promising result with 36 templates was found using MaSp1 amino acid sequences and 37 templates were found using MaSp2 amino acid sequences. However, the MaSp1 and MaSp2 templates were filtered based on target-template alignment and the templates with the highest quality prediction were chosen to undergo predicted protein model building assembly using ProMod3. The final models were constructed based on the target-template alignment using ProMod3 with the coordinates located between the target and the template was copied from the template to the model. Also, the model structure can be reconstructed back using fragment library through a process called insertions and deletions before rebuilding the side chains. Finally, the Qualitative Model Energy ANalysis (QMEAN) and the Global Model Quality Estimation (GMQE) score was used to assess the pre-residue model quality and estimate the accuracy of tertiary structure of the resulted model.

3.4 RaptorX Structural Elucidation of MaSp Proteins

RaptorX is an in silico program that predicts the structural properties of a protein sequence without the use of templates as mentioned by Wang *et al.,* [24]. This open-source platform from http://raptorx.uchicago.edu are known to excel in modelling servers for protein without close homologs in PDB or limited and far between sequence profile (i.e., had little evolutionary information) by exploiting structure information. The advantages of this server are that it's implementing a powerful in-house deep learning model DeepCNF (Deep Convolutional Neural Fields) which was used to predict the local structural properties such as the 3/8-state protein secondary structure (SS3/SS8), solvent accessibility (ACC) and disorder regions (DISO) as stated by Wang *et al.,* [24]. After the sequence was submitted the complete reported data was obtained from the RaptorX provided an insight into the predicted protein structure model based on protein sequence and the predicted secondary structure. The structural elucidation and figures obtained from RaptorX protein prediction algorithm for MaSps 1 and 2 were analysed thoroughly and illustrated in Figure 6 and Figure 7.

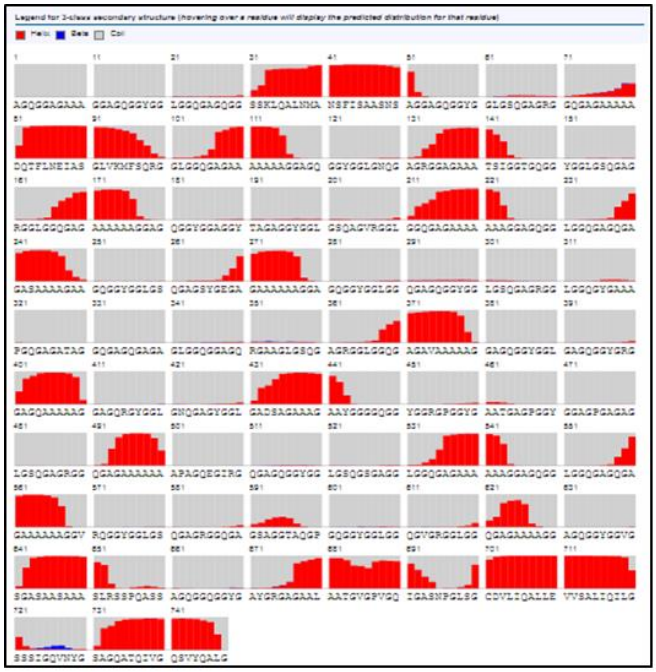

Fig. 6. Final model prediction of Major Ampullate Spidroin 1 (MaSp1) and 2 (MaSp 2) by Phyre2

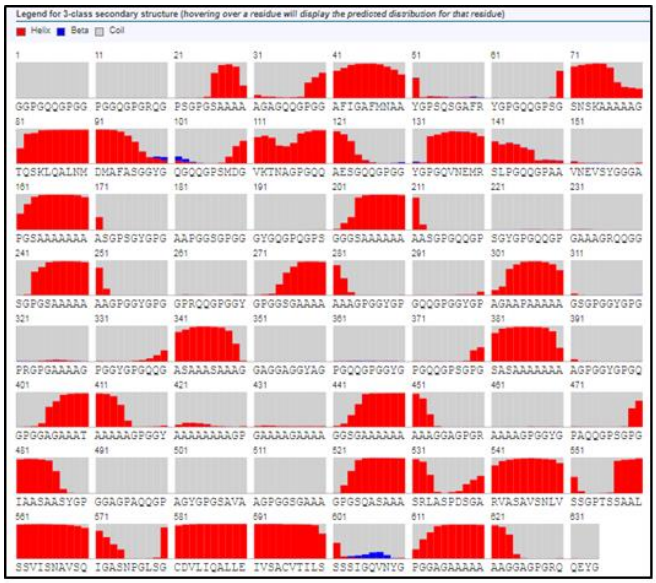

Fig. 7. Final model prediction of Major Ampullate Spidroin 1 (MaSp1) and 2 (MaSp2) by Phyre2

The distribution and local structure for the predicted secondary structure in percentage whereby each residue is colour coded either in red, blue or grey. When the residue has a high percentage of helix structure, it represents in red. If the residue has a high percentage in beta formation, it is represented in blue. Additionally, if the residue has a high percentage of coils formation it's represented in grey. The percentage of either helix, beta or coil in each residue can be seen through the indication bar colour (red, blue or grey) that provide better accuracy of each structure predicted in the secondary structure as emphasized by Wang *et al.,* [24]. Based on the predicted secondary structure in MaSp1 and MaSp2, the predicted protein structure model was constructed with the root-mean-square deviation (RMSD) value of 12.646 Å and 11.564 Å in MaSp1 and MaSp2, respectively.

3.5 Visualisation of Predicted 3-Dimensional (3D) Structure Models using UCSF Chimera

The UCSF Chimera is an extensible program for interactive visualisation and analysis of molecular structures and related data, including density maps, supramolecular assemblies, sequence alignments, docking results, trajectories, and conformational ensembles as described by Yuan *et al.,* [25]. Altogether, the data obtained from the protein threading analysis and multiple algorithms for structural prediction of MaSp1 and MaSp2 were collectively analysed, built and visualized using UCSF chimera software (Figure 8). The predicted 3D protein structure from all computational modelling algortihms (Phyre2, SWISS-MODEL and RaptorX) were visualised in UCSF Chimera to provide a more in-depth visualization and structural observation of MaSp1 and MaSp2's repetitive region models. The similarities and differences of utilizing each and different software were tabulated and explained in Table 1.

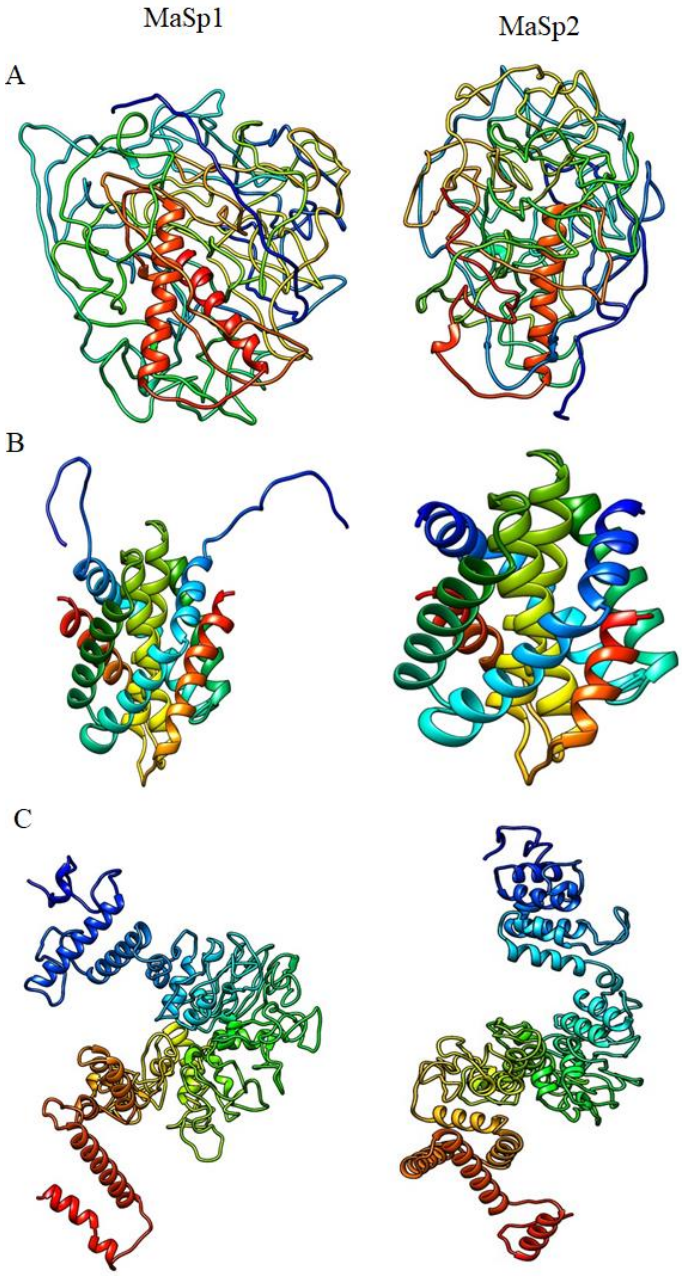

Fig. 8. The visualisation of the predicted MaSp1 and MaSp2's repetitive region models from the A) Phyre2, B: SWISS-MODEL, and C: RaptorX server using UCSF Chimera

Table 1
Similarities and differences of all computational approaches for Masp1 and 2 structural predictions

Criteria	Phyre2	SWISS-MODEL	RaptorX
Accuracy	Average 3-state accuracy of 75-80%	Accurate in 3D protein structure models	84% Q3[d] accuracy and 72% Q8[d] accuracy
Images of the predicted models	+	+	+
Predicted secondary structure	+	-	+
Predicted solvent accessibility	-	-	+
An evolutionary classification of proteins	+	-	-
Contact and distance map	-	-	+
Domain prediction and alignment	+	+	-
Predict Ligand binding site	+	+	+
Score	Confidence score	QMEAN[b] GMQE[c]	RMSD[a]
GMQE	RMSD[a]	SWISS-MODEL Templates Library (SMTL)	No templates used
Templates	Template-based ('homology') modelling	12	13
Rank CASP[g]	8		

Notes;

a) RMSD is the root mean square deviation
b) QMEAN is Qualitative Model Energy Analysis
c) GMQE is Global Model Quality Estimate
d) Q3 and Q8 is the percentages of residues for which the predicted secondary structures were correct
e) Z-score is the standard deviation from the mean
f) TM-Score is Template modelling score
g) CASP is Critical Assessment of protein Structure Prediction

4. Conclusions

Throughout this study, the three-dimensional (3D) protein models prediction of *N. pilipes* spider silk protein (MaSp1 and MaSp2) described and elucidated the *N. pilipes* spider silk protein (MaSp1 and MaSp2). The repeating region of the amino acid sequence using PDB and BLAST detected around 748 aa and 627 aa MaSp1 and MaSp2 of *N. pilipes* spidroin, respectively. The MaSp1 and MaSp2 amino acid sequences were converted to FASTA format and submitted to three different computational algorithms: Phyre2, SWISS-MODEL, and RaptorX. The UCSF Chimera software was used to visualise the projected 3D models created by these techniques to better gain an understanding of MaSp1 and MaSp2's repeating region. Finally, using UCSF Chimera, the Phyre2, SWISS-MODEL, and RaptorX servers developed the best-suited 3D structural models of MaSp1 and MaSp2 molecules. These 3D models' predictions of MaSp1 and MaSp2 will become beneficial in the field of protein engineering studies involving recombinant spider dragline silk proteins from *N. pilipes*, which will be a valuable asset for the future development of spider silk technology in Malaysia.

Acknowledgement
We would like to thank the Ministry of Higher Education, Malaysia for supporting this work through the Fundamental Research Grant Scheme (FRGS) Phase 1/2017 (9003 – 00631) for sponsoring this research and also the Faculty of Chemical Engineering Technology, UNIMAP, Perlis, and the Department of Analytical Biochemistry Research Centre (ABrC) USM, Penang for lab work accommodation.

References

[1] Whittall, Dominic R., Katherine V. Baker, Rainer Breitling, and Eriko Takano. "Host systems for the production of recombinant spider silk." *Trends in Biotechnology* 39, no. 6 (2021): 560-573. https://doi.org/10.1016/j.tibtech.2020.09.007
[2] Kiew, Peck Loo, and Jin Fung Toong. "Screening of significant parameters affecting Zn (II) adsorption by chemically treated watermelon rind." *Progress in Energy and Environment* 6 (2018): 19-32.
[3] Hajer, J., and Dana Řeháková. "Mating behavior of Theridiosoma gemmosum (Araneae: Theridiosomatidae): The unusual role of the male dragline silk." *Archives of Biological Sciences* 63, no. 1 (2011): 199-208. https://doi.org/10.2298/ABS1101199H
[4] Fai, Soon Zheng, and Choon Yoong Cheok. "Physical and chemical characterization of oil extracted from Citrofortunella microcarpa, Hibiscus sabdariffa and Artocarpus heterophyllus seeds." *Progress in Energy and Environment* 22 (2022): 1-12. https://doi.org/10.37934/progee.22.1.112
[5] dos Santos-Pinto, José Roberto Aparecido, Helen Andrade Arcuri, Helga Priewalder, Heliana Clara Salles, Mario Sergio Palma, and Gert Lubec. "Structural model for the spider silk protein spidroin-1." *Journal of Proteome Research* 14, no. 9 (2015): 3859-3870. https://doi.org/10.1021/acs.jproteome.5b00243
[6] Abdullah, N. A. H., J. A. Mohtar, KS Ku Ismail, and B. Y. Khoo. "A comparison of protein extraction methods suitable for gel-based proteomic studies of spider silk proteins." In *AIP Conference Proceedings*, vol. 2339, no. 1, p. 020129. AIP Publishing LLC, 2021. https://doi.org/10.1063/5.0046063
[7] Rim, Nae-Gyune, Erin G. Roberts, Davoud Ebrahimi, Nina Dinjaski, Matthew M. Jacobsen, Zaira Martín-Moldes, Markus J. Buehler, David L. Kaplan, and Joyce Y. Wong. "Predicting silk fiber mechanical properties through multiscale simulation and protein design." *ACS biomaterials science & engineering* 3, no. 8 (2017): 1542-1556. https://doi.org/10.1021/acsbiomaterials.7b00292
[8] Abu Bakar, Amirul Ridzuan, Thamilvaani Manaharan, Amir Feisal Merican, and Saharuddin Bin Mohamad. "Experimental and computational approaches to reveal the potential of Ficus deltoidea leaves extract as α-amylase inhibitor." *Natural Product Research* 32, no. 4 (2018): 473-476. https://doi.org/10.1080/14786419.2017.1312393
[9] Vakser, Ilya A. "Protein-protein docking: From interaction to interactome." *Biophysical journal* 107, no. 8 (2014): 1785-1793. https://doi.org/10.1016/j.bpj.2014.08.033
[10] Soni, Neelesh, and M. S. Madhusudhan. "Computational modeling of protein assemblies." *Current opinion in structural biology* 44 (2017): 179-189. https://doi.org/10.1016/j.sbi.2017.04.006
[11] Braitbard, Merav, Dina Schneidman-Duhovny, and Nir Kalisman. "Integrative structure modeling: overview and assessment." *Annual review of biochemistry* 88 (2019): 113-135. https://doi.org/10.1146/annurev-biochem-013118-111429
[12] Webb, Benjamin, and Andrej Sali. "Comparative protein structure modeling using MODELLER." *Current protocols in bioinformatics* 54, no. 1 (2016): 5-6. https://doi.org/10.1002/cpbi.3
[13] Christiansen, Tom, Larry Wall, and Jon Orwant. *Programming Perl: Unmatched power for text processing and scripting*. " O'Reilly Media, Inc.", 2012.
[14] De Smedt, Tom, and Walter Daelemans. "Pattern for python." *The Journal of Machine Learning Research* 13, no. 1 (2012): 2063-2067.
[15] Chamberlain, Scott A., and Carl Boettiger. *R Python, and Ruby clients for GBIF species occurrence data*. No. e3304v1. PeerJ Preprints, 2017. https://doi.org/10.7287/peerj.preprints.3304v1
[16] Cymerman, Iwona A., Marcin Feder, Marcin Pawłowski, Michal A. Kurowski, and Janusz M. Bujnicki. "Computational methods for protein structure prediction and fold recognition." *Practical bioinformatics* (2004): 1-21. https://doi.org/10.1007/978-3-540-74268-5_1
[17] Chan, Alexander. "An analysis of pairwise sequence alignment algorithm complexities: Needleman-wunsch, smith-waterman, fasta, blast and gapped blast." *Biochemistry–Final Project* (2007).
[18] Gundry, Rebekah L., Melanie Y. White, Christopher I. Murray, Lesley A. Kane, Qin Fu, Brian A. Stanley, and Jennifer E. Van Eyk. "Preparation of proteins and peptides for mass spectrometry analysis in a bottom-up proteomics

workflow." *Current protocols in molecular biology* 90, no. 1 (2010): 10-25. https://doi.org/10.1002/0471142727.mb1025s88

[19] Fiser, Andras. "Template-based protein structure modeling." *Computational biology* (2010): 73-94. https://doi.org/10.1007/978-1-60761-842-3_6

[20] 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

[21] Waterhouse, Andrew, Martino Bertoni, Stefan Bienert, Gabriel Studer, Gerardo Tauriello, Rafal Gumienny, 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

[22] Vouzis, Panagiotis D., and Nikolaos V. Sahinidis. "GPU-BLAST: using graphics processors to accelerate protein sequence alignment." *Bioinformatics* 27, no. 2 (2011): 182-188. https://doi.org/10.1093/bioinformatics/btq644

[23] Steinegger, Martin, Markus Meier, Milot Mirdita, Harald Vöhringer, Stephan J. Haunsberger, and Johannes Söding. "HH-suite3 for fast remote homology detection and deep protein annotation." *BMC bioinformatics* 20, no. 1 (2019): 1-15. https://doi.org/10.1186/s12859-019-3019-7

[24] Wang, Sheng, Wei Li, Shiwang Liu, and Jinbo Xu. "RaptorX-Property: a web server for protein structure property prediction." *Nucleic acids research* 44, no. W1 (2016): W430-W435. https://doi.org/10.1093/nar/gkw306

[25] Yuan, Shuguang, HC Stephen Chan, and Zhenquan Hu. "Implementing WebGL and HTML5 in macromolecular visualization and modern computer-aided drug design." *Trends in biotechnology* 35, no. 6 (2017): 559-571. https://doi.org/10.1016/j.tibtech.2017.03.009