

In-Silico Approaches for Molecular Characterization, Structural Function Prediction and Peptide Toxicity Analysis of the Matrix Protein of Vesicular stomatitis Indiana Virus (VSIV)

Ramya S*, Shri Varshini R, Poojaa Sri R, and Puniethaa Prabhu

Department of Biotechnology, K. S. Rangasamy College of Technology, Tiruchengode, Tamil Nadu, India.

Abstract. Vesicular Stomatitis Indiana Virus (VSIV) is a primary RNA virus studied for its effects in virology and in biotechnology the matrix protein of VSIV is key in linking the viral envelope to the core which is essential role in the assembly and release of viral particles. Through In silico approach the physicochemical properties secondary and tertiary structures and functional annotations of the matrix protein are analyzed, results indicate that the VSIV matrix protein is dominated by alpha-helical regions with a consistent tertiary structure validated by quality assessment methods. Peptide toxicity analysis of the matrix protein of VSIV focuses on evaluating the potential toxic effects of peptides derived from the matrix protein. Functional analysis confirms the protein role in forming virus particle. Molecular dynamic simulation to model the interaction of atoms and molecules and understand the behavior and properties of molecular structure in various condition.

Keywords: Vesicular Stomatitis Indiana Virus (VSIV), Matrix Protein, Molecular Characterization with Structure Prediction, Peptide Toxicity Analysis, Molecular Dynamic Simulation.

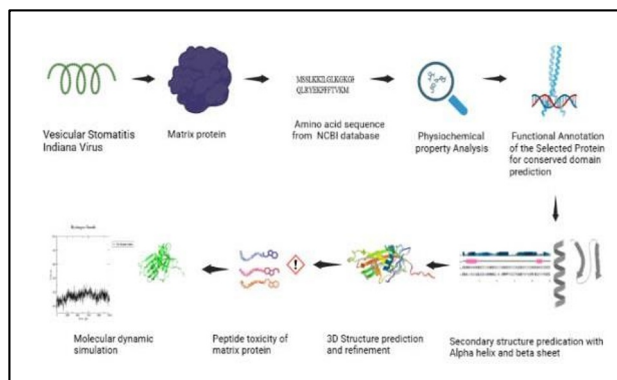


Fig. Graphical abstract of overall steps of the proposed system.

* Corresponding author: ramyasivasenthil@gmail.com

1 Introduction

Vesicular stomatitis Indiana virus (VSIV) is a vector-borne virus that can infects livestock and also affect human. VSIV was initially found in the year of 1998 in Peninsular Malaysia and identified in pigs and pig's farmers during an outbreak [1]. It is a zoonotic disease it can spread from animals to humans; human infections are usually mild [2]. VSIV classified within the Rhabdoviridae family and Vesiculovirus genus, which includes the New Jersey strain. The VSIV genome consists of a non-segmented, negative-sense, single-stranded RNA that encodes five proteins: matrix protein, nucleoprotein, glycoprotein, and phosphoprotein. Outbreaks of VSIV are seen in regions with high livestock populations in the US and Central America [3]. The matrix protein of VSIV plays a vital role in VSIV infectivity and structural integrity which makes it highly favourable for bioinformatics analysis. Bioinformatics tool can be used for analysis of amino acid sequence of matrix protein and to analysis conserved regions to the function of the molecular characterization. Structural prediction methods, Homology modelling are used to determine the 3D structure of matrix protein [4]. Analysis peptide toxicity, it provides insights into immunogenic properties of the protein. This analysis is important for development of safe and effective treatments [5]. The MD simulations are intended to mimic real physical movements of atoms and molecules in a protein structure over time. Such simulations can reveal the stability of the protein, changes in conformation, and the ability of the protein to interact with other molecules [6]. The ensuing sections explain the methodologies and tools used at each step of the analysis from the selection of protein to molecular dynamic simulations.

2 Materials and Methods

2.1 Selection of Protein and Sequence Obtaining

The amino acid sequences in FASTA format from the NCBI database of matrix protein found in VSIV was obtained [7]. This shows amino acid sequence information and helps to model the tertiary structure with functional characterization.

2.2 Characterization of the Physiochemical Properties of the VSIV Matrix Protein

The physiochemical properties of amino acid sequence composition, Theoretical isoelectric point (pI), Instability index, Aliphatic index, GRAVY and Estimated half-life of AHK23712.1 were obtained from ExPASy server ProtParam tool (web.expasy.org) [8].

2.3 Functional Prediction of the VSIV Matrix Protein

In the matrix protein (AHK23712.1), the NCBI-CDD server tool was predicted the conserved domain. Server containing conserved protein domain superfamily matrix of the VSIV matrix protein [9].

2.4 Assessment and Analysis of Secondary Structural Properties

The secondary structure elements of VSIV matrix protein were predicted by self-optimized prediction method with alignment (SOPMA) [10]. SOPMA analyses the sequence and provide secondary structure elements such as alpha- helices, beta-sheet and random coils. The Secondary structure was predicted using the PSIPRED (v.4.0) algorithm [11].

2.5 Tertiary Structural Prediction and Validation of the Matrix Protein

SWISSMODEL predicted the 3D structure and generates a predicted 3D model of the protein based on templates. Evaluation of model quality [1 2] . Ramachandran plot of model structure validated by PROCHECK tool of the SAVES (v.6.0) [13]. Galaxy WEB tool refines the model by optimizing its atomic interaction and accuracy of the predicted structure from matrix protein [14].

2.6 Toxicity Assessment of the Selected Protein

In silico hydrolysis of Matrix protein to peptide was predicted by BIOPEP-UWM tool. Toxicity of the peptide sequence of selected protein was predicted using the ToxinPred tool [15].

2.7 Molecular Dynamic Simulation of the Refined 3D Structure

Molecular dynamic simulation of refined tertiary structure that identifies the properties like RMSD, Gyration, Hydrogen Bond and predicting simulated 3D structure of the selected protein. The structural changes that occurred during the MD simulation for analysing these properties using WEBGRO tool [16].

3 Results and Discussion

3.1 Results

3.1.1 Protein and Sequence Obtaining

The amino acid sequence of the VSIV matrix protein AHK23712.1 was retrieved from the NCBI database. The protein sequence consisting of 230 amino acid was use to model the tertiary structure of protein. The protein information about locus is AHK23712, 230 Amino acid, Matrix protein; partial [Vesicular stomatitis Indiana virus] is a Definition, AHK2371 Accession number, version AHK23712.1, Source and Organism of VSIV matrix protein are both Vesicular stomatitis Indiana virus.

FASTA SEQUENCE: >AHK23712.1 M protein, partial [Vesicular stomatitis Indiana virus]
 MSSLKKILGLKGGKSKKLGIAPPPYEEDTNMEYAPSAPIDKSYFGVDEMDTHDPN
 QLRYEKFFFTVKMTVRSNRPFRTYSDVAAA VSHWDHMYIGMAGKRPFYKILAF LGS
 SNLKATPAVLADQGGQPEYHAHCEGRAYLPHRMGKTTPMLNVPEHFRFPFNIGLYKG
 TIELTMTIYDDESLEAAPMIWDHFNSKFSDFREKALMFGGLIVEKKASGAWVLDSVS
 HFKG

3.1.2 Identification of the Physicochemical Properties of the VSIV Matrix Protein

The amino acid sequence of the Matrix protein AHK23712.1 found in VSIV was analyzed for its physicochemical parameter. The [Table 1] represent the stability with an instability index of 38.18, as values be 40.00 indicate stability. The Theoretical pI of the protein is approximately 9.08, indicating that it is basic. The molecular weight, aliphatic index, instability index, and GRAVY index then the values are as follows 26108.02, 66.22, 38.18, and -0.455, respectively. The aliphatic index value of 66.22 indicates moderate thermo stability, indicating that the protein remains stable over standard temperature ranges of 20-60 degrees Celsius, the temperature ranges is advantageous. The Aliphatic Index value of is 66.22 indicates the protein has high thermo-stability, allowing it to remain stable over a

wide temperature range. GRAVY Index value is -0.455 indicates the protein is hydrophilic interaction.

Table 1. Physiochemical Properties.

Physiochemical parameters	Values
Number of amino acids	230
Molecular weight	26108.02
Theoretical isoelectric point pI	9.08
Total number of negative charged residues (Asp + Glu)	26
Total number of positive charged residues (Arg + Lys)	31
Formula	C1185H1810N312O331S12
Total number of atoms	3650
The estimated half-life	30 hours (mammalian reticulocytes, in vitro). >20 hours (yeast, in vivo). >10 hours (E. coli, in vivo)
Instability index	38.18
Aliphatic index	66.22
Grand average of hydropathicity (GRAVY)	-0.455

3.1.3 Functional Annotation of the VSIV Matrix Protein

Protein AHK23712.1 has been identified as a matrix protein through analysis using the NCBI Conserved Domain Database (CDD). The conserved domain was located at positions 23–222 (e-value of 4.90e-121) and the actual alignment was detected with superfamily member pfam06326. The lone member of the superfamily cl05690. The CDD [fig. 1] that identifies viral matrix protein domains in the query sequence by comparing a conserved protein domain family matrix.

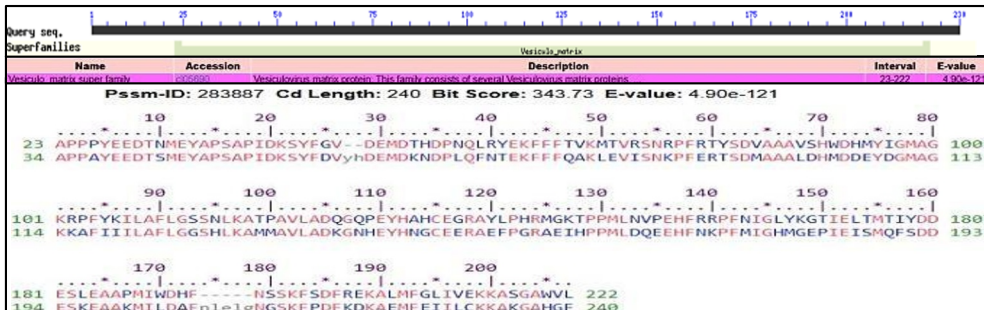


Fig. 1. Functional prediction of the VSIV matrix Protein (AHK23712.1.) the graphical figure represents the identified in the sequence. An * (asterisk) mentioned positions with a single, completely conserved residue.

3.1.4 Secondary Structural Properties

The interaction between structure and function in proteins. For determining the function, structure, of a protein, secondary structural features such as helices, coils, sheets, or turns are of significance. The secondary structural components of protein AHK23712.1 are predicted using the SOPMA software. The [Table 2] represents amino acid composition and [Table 3] Indicated about the Alpha Helix (Hh) were 62 is 26.96%, Extended Strand (Ee) were 45 is 19.57%, Beta Turn (Tt) were 0 is 0.00% Random Coil (Cc) were 123 is 53.48% with greater confidence. The [fig. 2] labeled the secondary structure using PSIPRED software.

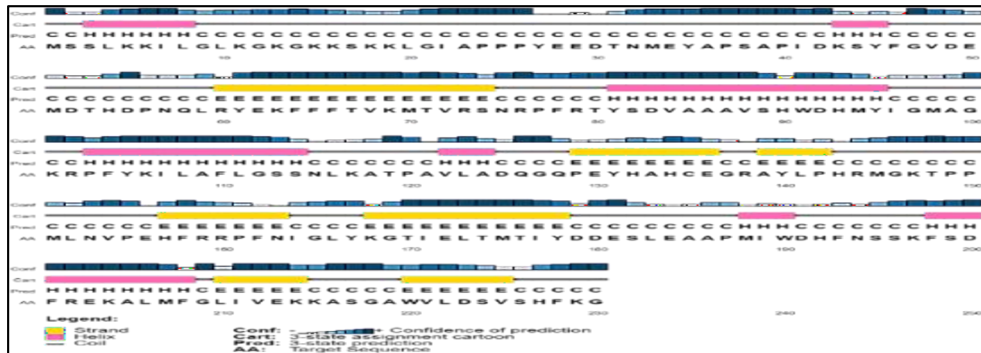


Fig. 2. Secondary structure of the VSIV Matrix protein.

Table 2. Amino acid composition.

Amino Acids	Percentage (%)
Ala (A)	7.8%
Arg (R)	4.3%
Asn (N)	3.0%
Asp (D)	5.7%
Cys (C)	0.4%
Gln (Q)	1.3%
Glu (E)	5.7%
Gly (G)	7.0%
His (H)	3.9%
Ile (I)	4.3%
Leu (L)	7.4%
Lys (K)	9.1%

Table 3. Secondary Structural Elements.

Secondary Structural Elements	Values (%)
Alpha helix (Hh)	62(26.96%)
3 ₁₀ helix (Gg)	0 (0.00%)
Pi helix (Ii)	0 (0.00%)
Beta bridge (Bb)	0 (0.00%)
Extended strand (Ee)	45(19.57%)
Beta turn (Tt)	0 (0.00%)
Bend region (Ss)	0 (0.00%)
Random coil (Cc)	123(53.48%)
Other states	0 (0.00%)

3.1.5 Prediction and Validation the Tertiary Structural of the Matrix Protein

The tertiary structure of the protein was then predicted using the Swiss Model in the [fig. 3]. The [fig. 3 (a)] and [fig. 3 (b)] highlighted the QMEAN score is -0.81 and ERRAT value. Sequence identity 96.26% to assess the quality of the modelled structure. [Fig. 3 (c)] display the Ramachandran plot generated by PROCHECK that 89.2% of the residues were located in the core regions (A, B, L), indicating that they adopt favorable conformations. Only 10.8% of the residues were in the additional allowed regions (a, b, l, p), and 0.0% were in the generously allowed regions, the structure's stability and accuracy. The analysis also reported the numbers of glycine and proline residues (10 and 11, respectively) and the total non-glycine and non-proline residues (158). The total number of residues is 181. ERRAT value of 98.113% were in the overall quality factor [17]. The Verify 3D tool confirmed that the predicted tertiary structure passed the evaluation and confirmed its reliability for further functional and structural analyses. Structural Refinement for Optimizing side-chain conformations that refers to the adjustment of amino acid side chains to assume the most energetically favorable configuration, were RMSD 0.393. Refined model in the [fig. 3 (d)]. It is an important step in refining the overall structure of the protein, to accurately model how the protein interacts with other molecules using the Galaxy WEB tool. The Ramachandran plot generated by PROCHECK showed in the [fig. 3 (e)] that 96.8% of the residues were located in the core regions (A, B, L), indicating that they adopt favorable conformations. Only 3.2% of the residues were in the additional allowed regions (a, b, l, p), and 0.0% were in the generously allowed regions, the structure's stability and accuracy. The analysis also reported the numbers of glycine and proline residues (10 and 11, respectively), the total non-glycine and non-proline residues (158) and the total number of residues is 181.

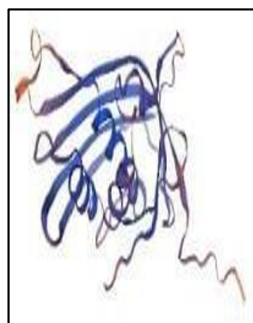


Fig. 3. Tertiary structure was predicted using the Swiss Model.

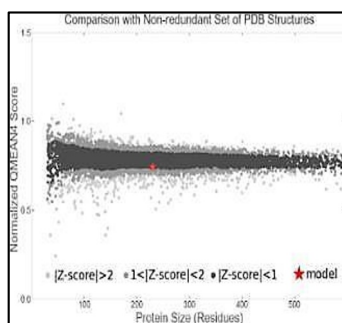


Fig. 3 (a) Quality analysis with -0.81 QMEAN score.

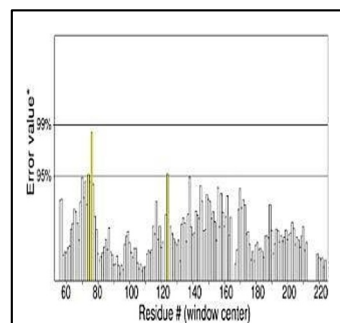


Fig. 3 (b) ERRAT value of 98.113% with high resolution.

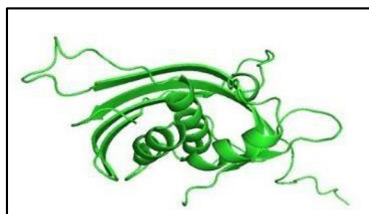


Fig. 3 (d) Structural refinement using Galaxy WEB.

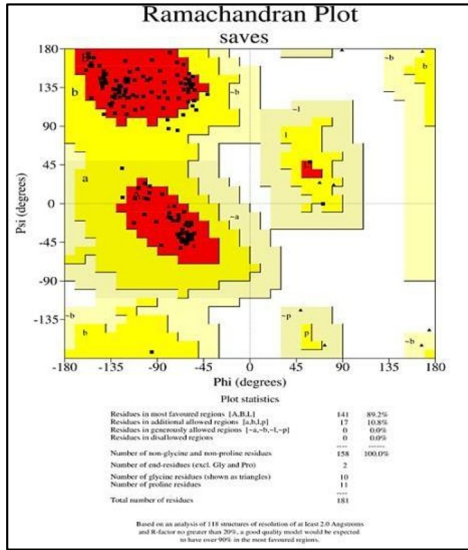


Fig. 3 (c) Ramachandran plot rendered by SAVES (v.6.0) with 89.2 % favoured regions.

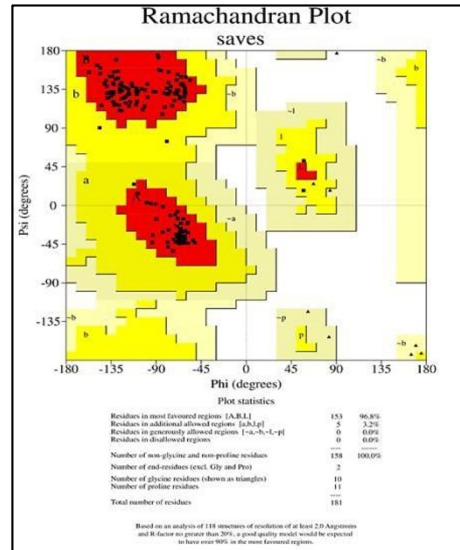


Fig. 3 (e) Ramachandran plot with 96.8% favoured regions after refinement.

3.1.6 Toxicity Prediction of the Selected Protein

In silico hydrolysis of the VSIV matrix protein into peptide sequences using the ToxinPred and BIOPEP-UWM tools with enzymes pepsin and trypsin. Pepsin degrades proteins into smaller peptides under acidic conditions, while trypsin further degrades those peptides into amino acids in neutral conditions. These were used for toxicity prediction and the exploration of bioactivity, biased more on peptides with five or more amino acids. The peptide sequences of pepsin and trypsin give below. The [fig. 4] illustrated the peptide toxicity of the given protein.

Trypsin Peptide sequence: MSSLK-K-ILGLK-GK-GK-K-SK-K-LGIAPPPYEEDTNMEYAPSAPIDK-SYFGVDEMDTHDPNQLR-YEK-FFFTVK-M TVR -SNR-PFR TYSDVAAAVSHWDHMYIGMAGK-R-PFYK-ILAFGLSSNLK-ATPAVLADQQPEYHAHCEGR-AY LPHR - MGK - TPPMLNVPEHFR- R - PFNIGLYK - GTIELTMTIYDDESLEAAPMIWDHFNSSK - FSDFR - EK - ALMFGL IVEK - K -ASGAWVLDSVSHFK – G

Pepsin peptide sequence: MSSL-KKIL-GL-KGKGKSKKL-GIAPPPYEEDTNMEYAPSAPIDKSYF-GVDEMDTHDPNQL-RYEKF- F - F - TVKM TVRSNRPF - RTYSDVAAAVSHWDHMYIGMAGKRPF- YKIL - AF - L - GSSNL - KATPAVL - ADQGQPEYHAHCEGRAY L - PHRMGKTPPML - NVPEHF -RRPF - NIGL - YKGTIEL - TMTIYDDES L - EAAPMIWDHF - NSSKF - SDF - REKAL - MF - GL - IVEKKASGAWVL - DSVSHF – KG

Peptide ID	Peptide Sequence	SVM Score	Prediction	Hydrophobicity	Hydropathicity	Hydrophilicity	Charge	Mol wt
	MSSLK	-0.98	Non-Toxin	-0.17	0.04	0.10	1.00	564.76

Fig. 4. Shows non-toxin prediction with SVM score value is -0.98, molecular weight of 564.76, -0.17 hydrophobicity indicates slightly hydrophilic character, 0.04 hydrophaticity indicates slightly hydrophobic character, 0.10 hydrophilicity indicates moderate hydrophilic character of the trypsin peptide sequence of the selected protein. Both the peptide sequence is resulting in non-toxin.

3.1.7 Performing Molecular Dynamic Simulation of the Optimized 3D Structure

Molecular dynamic simulation of the refined 3D structure analyzing by WEBGRO tool visualize in [fig. 5]. The WEBGRO tool using in this simulation that identifies properties like [fig. 5 (a)] shows value of RMSD is 0.3nm (3Å), which is well folded and stable with minor fluctuations and [fig. 5 (b)] shows Gyration values around 1.65nm imply a stable, folded structure and [fig. 5 (c)] shows Hydrogen bond values around 120 indicate stable interactions, which are analyzed within the molecular structure of the VSIV matrix protein. Molecular dynamic simulation helps study the stability, interaction, and conformational changes of molecules under various situations.

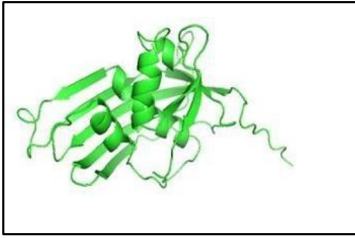


Fig. 5. Molecular dynamic simulation of enhanced 3D structure of the matrix protein.

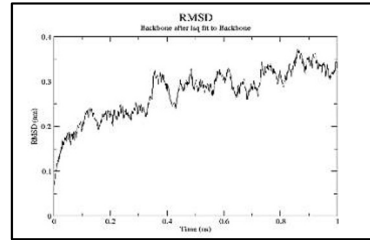


Fig. 5 (a) RMSD query for MD simulation using WEBGRO.

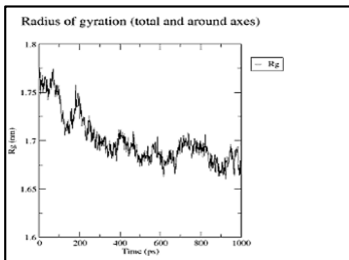


Fig. 5 (b) Gyration query for MD simulation using WEBGRO.

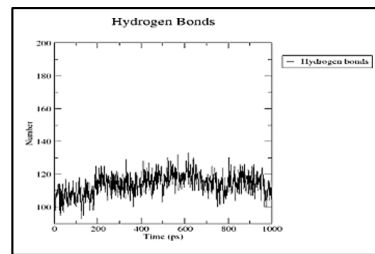


Fig. 5 (c) Hydrogen Bond query for MD simulation using WEBGRO.

3.2 Discussion

In this paper, on the analysis of the VSIV matrix protein AHK23712.1 the Sequence, physicochemical properties, functional annotation, and structural features. The protein since it has an instability index of 38.18 and an extremely high theoretical isoelectric point of 9.08, indicating it to be very basic. Functional annotation with the help of NCBI Conserved Domain Database (CDD) and secondary structure prediction using SOPMA and PSIPRED tools. Model validation has done using Swiss Model and PROCHECK tools which provide authentic models with QMEAN score -0.81 with 96.26% sequence identity that indicates reliability of the model. Ramachandran plot showed stability, with the placing of 96.8% of the residues within the favorable region. Such structural optimization which allows an accurate readjustment of the side chains for an RMSD value of 0.393, There is no toxicity in a peptide sequence of the matrix protein confirmed by a non-toxicity prediction. Molecular dynamics was analyzed using WEBGRO.

4. Conclusion

In detail, understanding methodology on the VSIV matrix protein study shows a detailed approach for investigating and validating the structural and functional characteristics of that protein. Considering the choice to characterizing the protein sequence, physicochemical properties of the sequence are assumed. Functions that are the conserved domains can predict this by using algorithms on predicting secondary structure elements. It also gives a very high prediction and refinement of the tertiary structure along with proper explanation in this context. In-silico hydrolysis and toxicity prediction, shows its usage as a therapeutic medicine. Stability and interaction property calculation by molecular dynamics analysis about finalized 3D structure brings out the mechanism of actions of proteins. It explains mechanisms of action by the understanding of how the protein works, together with the process with which the protein is connected to the assembling of the virus. Generally, these methods indicate great prospects for computational techniques in the context of present-day biomedical research and development of a therapeutic approach.

Acknowledgement

This study does not involve human participants, animals, or sensitive data requiring ethical approval or consent to participate. The data and materials used in this study were obtained from freely available online tools and resources, and can be accessed through the respective web platforms. The authors declare that they have no competing interests. No specific funding was received for this research work.

List of Abbreviations

NCBI: National centre for Biotechnology information; FASTA: Fast-All; ProtParam: Protein Parameter Tool; GRAVY: Grand average of hydropathicity; NCBI-CDD: National Center for Biotechnology Information - Conserved Domain Databases; SOPMA: Self-Optimized Prediction Method with Alignment; PSIPRED: PSI-BLAST based secondary structure prediction; QMEAN: Qualitative Model Energy Analysis; Swiss Model: Swiss Institute of Bioinformatics Model; Galaxy WEB: Galaxy Web-based platform; BIOPEP-UWM: Bioinformatics Peptide University of Warmia and Mazury; ToxinPred: Toxin Prediction tool; RMSD: Root Mean Square Deviation; WEBGRO tool: Web-based Genome Research tool; PROCHECK: Protein Structure Validation tool; SAVESv6.0: Structural Analysis and Verification Server Version 6.0.

References

1. Miranda R. Bertram, Vesicular stomatitis Indiana virus near-full-length genome sequences reveal low genetic diversity during the 2019 outbreak in Colorado, USA. *Frontiers in Veterinary Science*, **10**, pp. 1-5, (2023). <https://doi.org/10.3389/fvets.2023.1110483>
2. Balázs Kakuk, Nanopore Assay Reveals Cell-Type-Dependent Gene Expression of Vesicular Stomatitis Indiana Virus and Differential Host Cell Response, *Pathogen MDPI*, **10**, pp. 1-12, (2021). <https://doi.org/10.3390/pathogens10091196>
3. Selene Zarate, Phylogenomic Signatures of a Lineage of Vesicular Stomatitis Indiana Virus Circulating During the 2019–2020 Epidemic in the US, *Virus*, **16**, pp.1-26, (2024). <https://doi.org/10.3390/v16111803>

4. Abu Saim Mohammad Saikat, In-Silico Approaches for Molecular Characterization and Structure-Based Functional Annotation of the Matrix Protein from Nipah henipavirus, chemistry proceeding **12(1)**, pp. 1-10, (2022). <https://doi.org/10.3390/ecsoc-26-13522>
5. Qiule Yu, Zhixing, ToxGIN: an In silico prediction model for peptide toxicity via graph isomorphism networks integrating peptide sequence and structure information, Briefings in Bioinformatics, **25**, pp. 1-11, (2024). <https://doi.org/10.1093/bib/bbae583>
6. Kishori Lal, Study of Molecular Interaction Between Arabidopsis Aquaporin and Pseudomonas syringae Harpin(HrpZPss) Through Molecular Docking Tools, International Journal of Plant and Environment, **9**, pp. 343-349, (2023). [doi:10.18811/ijpen.v9i04.06](https://doi.org/10.18811/ijpen.v9i04.06)
7. Eric W. Sayers, Database resources of the National Center for Biotechnology Information. Nucleic Acids Research, **9**, pp. 10-17, (2020). [doi:10.1093/nar/gkaa892](https://doi.org/10.1093/nar/gkaa892)
8. Elisabeth Gasteiger. Protein Identification and Analysis Tools on the ExPASy Server. pp. 571-607, (2005).
9. Shennan Lu. D1, CDD/SPARCLE: the conserved domain database in 2020, Nucleic Acids Research, **48**, pp. 265-268, (2019). [doi:10.1093/nar/gkz991](https://doi.org/10.1093/nar/gkz991)
10. Geourjon .C, Deléage .G., SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Bioinformatics, **11**, pp. 681-684, (1995). <https://doi.org/10.1093/bioinformatics/11.6.681>
11. McGuffin, Liam J., The PSIPRED protein structure prediction server. Bioinformatics, Bioinformatic, **16**, pp. 404-405, (1999). [doi:10.1093/bioinformatics/16.4.404](https://doi.org/10.1093/bioinformatics/16.4.404)
12. Schwede, Torsten, SWISS-MODEL: an automated protein homology-modeling server, Nucleic Acids Research, **31**, pp. 3381-3385, (2003). [doi:10.1093/nar/gkg520](https://doi.org/10.1093/nar/gkg520)
13. Anantapur R, The 3D structure prediction, quality assessment and model validation of Replicase protein of MYMV, The Pharma Innovation Journal, **11**, pp.516-521, (2022).
14. Ko, Junsu, GalaxyWEB server for protein structure prediction and refinement, Nucleic Acids Research, **40**, pp. 294-297, (2012). [doi:10.1093/nar/gks493](https://doi.org/10.1093/nar/gks493)
16. Tumskiy, Roman S., Multistep rational molecular design and combined docking for discovery of novel classes of inhibitors of SARS-CoV-2 main protease 3CLpro. pp. 1-8, (2021). <https://doi.org/10.1016/j.cplett.2021.138894>
17. Tran, Ngoc Tuan, In silico characterisation, homology modelling and structure-based functional annotation of blunt snout bream (Megalobrama amblycephala) Hsp70 and Hsc70 protein, Journal of Animal Science and Technology, pp.1-9, (2015). [doi:10.1186/s40781-015-0077-x](https://doi.org/10.1186/s40781-015-0077-x)