Research Article

Structure-based computational approaches for characterization and functional elucidation of the fusion protein from *Nipah henipavirus*

Abu Saim Mohammad Saikat[1,*]

[1]Department of Biochemistry and Molecular Biology, Life Science Faculty, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj 8100, Bangladesh; asmsaikat.bmb@gmail.com

Abstract

Many epidemics of varying severity have triggered panic and devastation in the past. The Nipah virus has one of the world's highest fatality rates. The encephalitis resulting from acute respiratory distress has been fatal in some instances. Many factors influence the virus's genesis and spread. Developing new methods has improved personal hygiene awareness and surveillance over the contaminated area. An unidentified protein from *Nipah henipavirus* was the focus of this investigation. The secondary structure of the protein consists of a helix, a sheet, a turn, and a coil. Furthermore, the Ramachandran plot and the Z-score-based and local model quality assessment processes revealed the quality of the modeled protein structure. The protein can be used as a target for developing prospective antiviral medication and vaccine candidates.

Keywords Nipah virus; Protein characterization; Protein stability; GRAVY, Viral infection; Protein-based drug design; Protein-based vaccine

1 Introduction

The Nipah virus (NiV), spread by bats and can cause fatal encephalitis in people, has recently been identified in Malaysia, Bangladesh, Singapore, and India [1-3]. It belongs to the order Mononegavirales, which contains other developing lethal zoonotic viruses, including Hendra, Marburg, and Ebola [4]. The virus is thought to be stored naturally in the bodies of Pteropus fruit bats. Humans got NiV from pigs, the intermediate hosts of the virus, in 1998 during the first documented epidemic in the Malaysian town of Sungai [5-7]. Since 2001, the intake of raw date palm sap contaminated with the saliva and excreta of the bats has been reported as the source of yearly NiV outbreaks in various districts of Bangladesh. The first epidemic in India was recorded in Siliguri, West Bengal, in 2001, and it was mainly spread by intimate personal contact or nosocomial transmission. In 2007, a second outbreak was reported in Nadia and West Bengal [7, 8]. In a recent NiV epidemic in the Kozhikode region of Kerala, a state in South India, the index patient was said to have been infected by fruit-eating bats [9]. While nosocomial transmission

accounted for the vast majority of cases, no clinical or statistical data was provided to confirm the frequency of the illness. The most recent epidemic in Kerala had a death rate of 91%, which is typical of all outbreaks [9, 10].

Cell-cell fusion (syncytia) in lung, brain, kidney, and heart tissues is caused by Nipah (NiV) and Hendra (HeV) viruses. This results in encephalitis, pneumonia, and frequent death. Henipavirus infections are characterized by membrane fusion, which is required for viral entry and virus-induced cell-cell fusion [11-14]. Understanding the pathobiology of henipaviruses relies on elucidating the mechanism(s) of membrane fusion, which may lead to discovering new approaches to creating antiviral therapeutics. Viral attachment (G) and fusion (F) glycoproteins must work together to facilitate membrane fusion in henipaviruses. Current theories of henipavirus fusion propose that F is released from its metastable pre-fusion conformation to promote membrane fusion after NiV or HeV G attachment to its cell surface receptors [11, 15-18]. The selected protein for this study is a fusion protein of *Nipah henipavirus* associated with viral infections. The physicochemical characteristics and anticipated protein structures of the selected protein demonstrated structure-function relationships of the proteins associated with viral infections. Therefore, this protein can be targeted for predicting antiviral drugs and vaccines against the selected protein to combat viral infections.

2 Materials and Methods

2.1 Protein sequence retrieval

The protein sequence (GenBank: QBQ56722.1, NCBI accession: QBQ56722) was retrieved in FASTA format from the NCBI protein sequence database [19].

2.2 Identification of the physicochemical properties

The physicochemical characteristics of the protein were demonstrated by using the ExPASy ProtParam tool [20] and the SMS (v.2.0) program [21].

2.3 Secondary structure identification and assessment of the selected protein

The SOPMA program [22] was used following the default parameters (output width = 8; the number of conformational states = 4; helix, sheet, turn, and coil; similarity threshold = 8, and window width = 17) to determine the secondary structural parameters. Moreover, the SPIPRED program (v.4.0) [23] was used the determination of the secondary features and topology of the selected protein.

2.4 Determination and validation of the three-dimensional protein structure

The three-dimensional structure of the selected protein was anticipated by using the Modeller [24] with HHpred interface [25, 26]. Moreover, the PROCHECK program of the SAVES program (v.6.0) [27] was used for the structural validation of the modeled 3D structure of the protein. Also, the ProSA-web program [28] was used to determine the Z-score of the modeled structure for structural assessment.

3 Results and discussion

3.1 Sequence retrieval of the selected protein

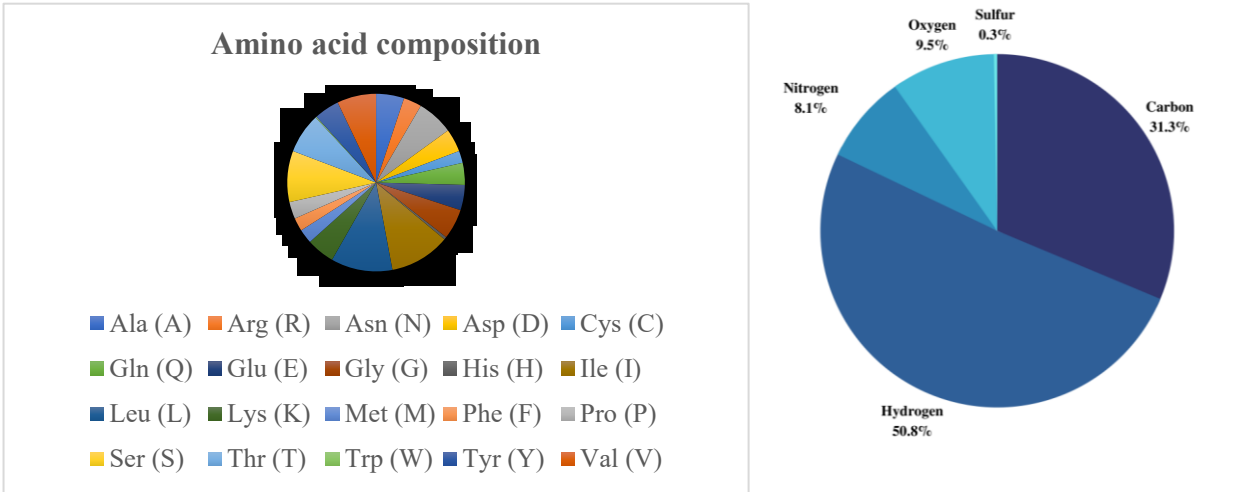
The protein sequence retrieved from the NCBI database contains 546 amino acid residues (Table 1). The fusion protein (accession no. QBQ56722, version no. QBQ56722.1) is found in the QBQ56722 locus of *Nipah henipavirus*.

Table 1. Protein retrieval

Protein individualities	Protein information
Locus	QBQ56722
Amino acid	546 aa
Accession	QBQ56722
Version	QBQ56722.1
GenBank ID	QBQ56722.1
Source	*Nipah henipavirus*
Organism	*Nipah henipavirus*
FASTA sequence	>QBQ56722.1 fusion protein [*Nipah henipavirus*] MAVILNKRYYSNLLLLILMISECSVGILHYEKLSKIGLVKGIT RKYKIKSNPLTKDIVIKMIPNVSNMSQCTGSVMENYKTRLNG ILTPIKGALEIYKNNTHDLVGDVRLAGVIMAGVAIGIATAAQI TAGVALYEAMKNADNINKLKSSIESTNEAVVKLQETAEKTV YVLTALQDYINTNLVPTIDKISCKQTELSLDLALSKYLSDLLF VFGPNLQDPVSNSMTIQAISQAFGGNYETLLRTLGYATEDFD DLLESDSITGQIIYVDLSGYYIIVRVYFPILTEIQQAYIQELLPV SFNNDNSEWISIVPNFILVRNTLISNIEIGFCLITKRSVICNQDY ATPMTNNMRECLTGSTEKCPRELVVSSHVPRFALSNGVLFA NCISVTCQCQTTGRAISQSGEQTLLMIDNTTCPTAVLGNVIISL GKYLGSVNYNSEGIAIGPPVFTDKVDISSQISSMNQSLQQSKD YIKEAQRLLDTVNPSLISMLSMIILYVLSIASLCIGLITFISFIIVE KKRNTYSRLEDRRVRPTSSGDLYYIGT

3.2 Physicochemical parameters determination of the selected protein

The physicochemical parameters of a protein are defined by the characteristics of its constituent amino acids. The alpha-carbon unit of all amino acids, except for glycine, is asymmetric, indicating that it is connected to four distinct chemical constituents (atoms or atom pairs) [29, 30]. Consequently, amino acids, except glycine, can appear in two distinct spatial or geometric configurations (i.e., isomers), which resemble left and right hands [31-33]. ExPASy ProtParam tool identified the physicochemical characteristics of the protein, such as amino acid compositions, atomic composition, and protein half-life calculation (Figure 1). Leucine is the most abundant amino acid (61, 11.2%) compared to others in the amino acid sequence. Moreover, the atomic composition of the protein demonstrated that hydrogen is the most abundant element (4361, 50.8%), following oxygen (817, 9.5%), nitrogen (693, 8.1%), and sulfur (26, 0.3%).



(a) (b)

Figure 1. Physicochemical parameters of the selected protein. (a) The protein contains Ala (28, 5.1%), Arg (18, 3.3%), Asn (36, 6.6%), Asp (23, 4.2%), Cys (12, 2.2%), Gln (22, 4.0%), Glu (25, 4.6%), Gly (30, 5.5%), His (3 , 0.5%), Ile (60, 11.0%), Leu (61, 11.2%), Lys (28, 5.1%), Met (14, 2.6%), Phe (13, 2.4%), Pro (17, 3.1%), Ser (51, 9.3%), Thr (40, 7.3%), Trp (1, 0.2%), Tyr (25, 4.6%), and Val (39, 7.1%). (b) The atomic composition of the protein as of carbon (2687, 31.3%), hydrogen (4361, 50.8%), nitrogen (693, 8.1%), oxygen (817, 9.5%), and sulfur (26, 0.3%).

The protein has a molecular weight of about 60280.90 Da (Table 2) with a theoretical pI of 6.08 (6.30*). The protein has the total number of positively charged residues (Arg + Lys), the whole number of atoms, and the absolute number of negatively charged residues (Asp + Glu) as of 46, 8584, and 48, respectively. As more protein therapies are being developed, many of which have a short plasma half-life, the biotech and pharmaceutical industries are focusing more and more on methods to lengthen that half-life [34, 35]. The therapeutic and cost benefits of a longer half-life are apparent. Numerous recognized or in-development biotherapeutics have a short half-life, needing numerous administrations to sustain a therapeutic level over a long period [36-38]. The use of half-life extension techniques permits the production of medicines with enhanced

pharmacokinetic and pharmacodynamic characteristics that have a prolonged half-life. Incorporating half-life extension methods into developing numerous biotherapeutics is now standard practice. Various options are available for fine-tuning half-life and adaptation to the desired treatment method and condition [39-42]. The anticipated protein half-life as of 30 hours (mammalian reticulocytes, *in vitro*); >20 hours (yeast, *in vivo*); and >10 hours (*Escherichia coli*, *in vivo*).

Efforts are undertaken to establish a relationship between the metabolic stability of proteins and aspects of their primary sequence and to use weight estimates of instability for a protein of established sequence to determine its resilience properties [43-46]. Proteins may be evaluated for viability *in vitro* using the 'Instability index.' If the index is under 40, the substance will likely be stable in the test tube. It is presumably not sustainable if it is more significant [47-49]. The instability index of the selected protein is 38.05 (less than 40.00), resulting in a stable nature. The aliphatic index measures how much space is taken up by a protein's aliphatic side chains compared to its total volume [50]. The thermal stability of proteins is related to their aliphatic index. Proteins with a high aliphatic index are less likely to denature when heated. Hydrophobicity is a property shared by aliphatic amino acids [50-52]. The aliphatic index of the selected protein is demonstrated as 112.27. GRAVY is the value employed to demonstrate a protein's hydrophobicity. This value is computed by accepting the absolute hydropathy values of all amino acids (aa) and splitting that whole by the entire sequence length [53-56]. The estimated GRAVY of the protein is 0.177.

Table 2. Physicochemical parameters of the selected protein

Parameters	Values
Molecular weight	60280.90 Da
Theoretical pI	6.08 (6.30*)
Total number of positively charged residues (Arg + Lys)	46
Total number of negatively charged residues (Asp + Glu)	48
Total number of atoms	8584
Estimated half-life	a) 30 hours (mammalian reticulocytes, *in vitro*) b) >20 hours (yeast, *in vivo*) c) >10 hours (*Escherichia coli*, *in vivo*)
Instability index (II)	38.05
Aliphatic index	112.27
Grand average of hydropathicity (GRAVY)	0.177

*pI calculated by the SMS v2.0 tool.

3.3 Identification and validation of the predicted secondary structure of the selected protein

In the context of a polypeptide chain, the term "secondary structure" refers to the standard and recurrent spatial configurations of neighboring amino acid residues. Hydrogen bonds between amide hydrogens as well as carbonyl oxygens of the peptide backbone are responsible for its stability. Alpha-helices (α-helices) and beta-structures (β-structures) are the two most important types of secondary structures [57-59]. The SOPMA program demonstrated that the protein contains alpha helix (239, 43.77%), extended strand (112, 0.51%), beta turn (23, 4.21%), and random coil (172, 31.50%). No Pi helix, beta bridge, bend region, and ambiguous states were present in the protein (Figure 2). The selected protein contains polar, non-polar, aromatic group-containing, and hydrophobic amino acid residues in its structure (Figure 3). Moreover, the sequence plot demonstrated the protein parameters, including the protein's helical, coil, and extracellular properties (Figure 3). The secondary structure of the selected protein is illustrated in Figure 4.

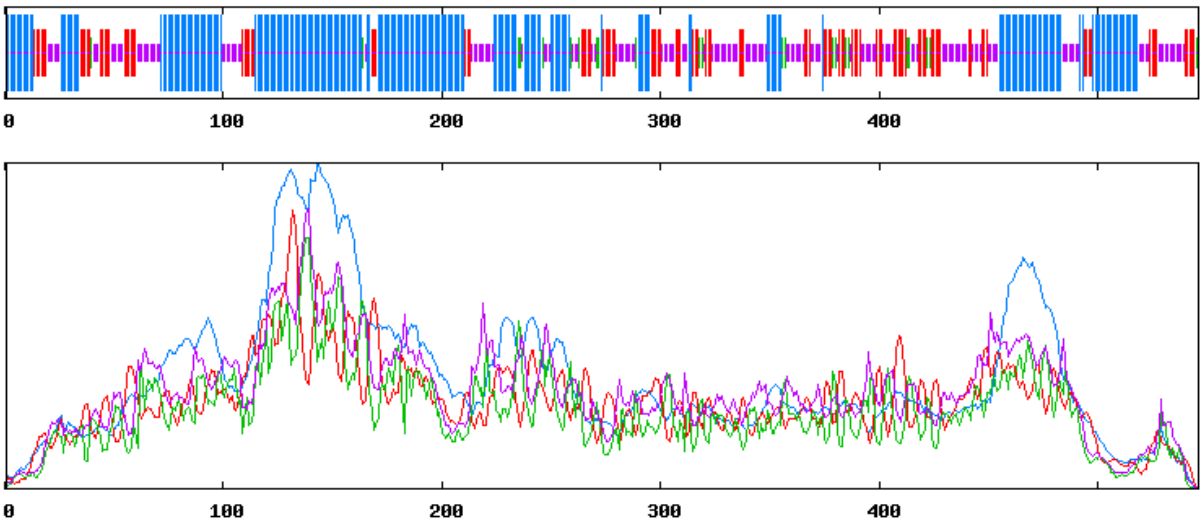
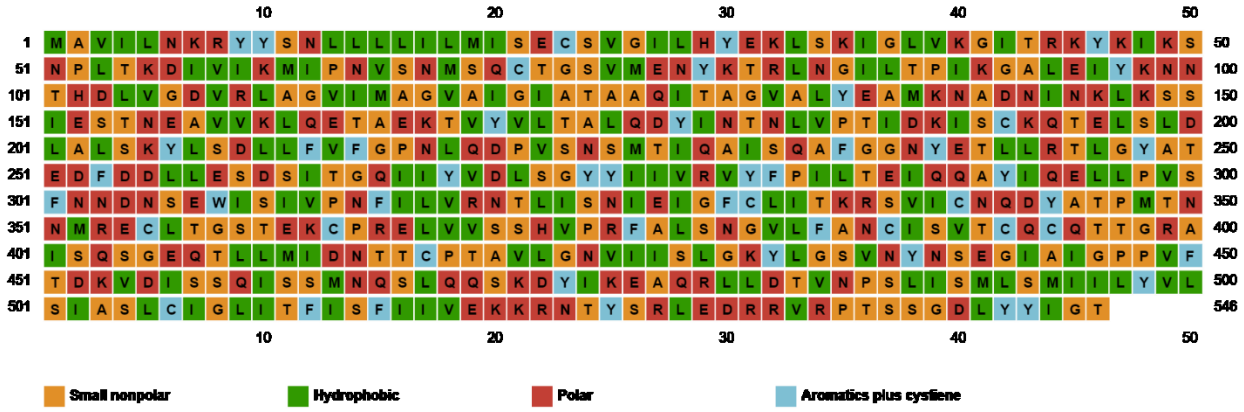
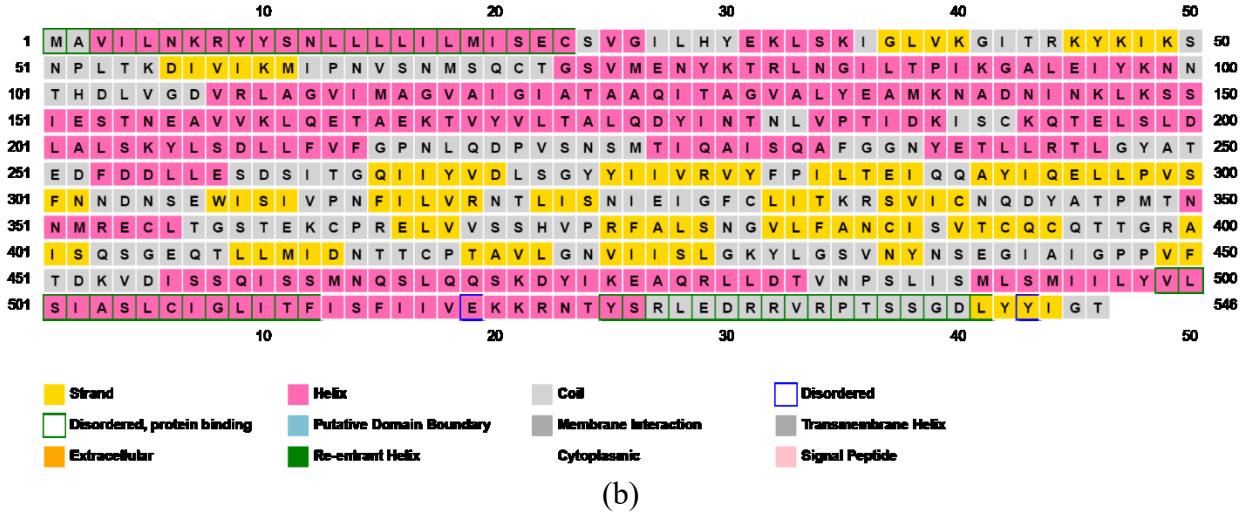


Figure 2. Secondary structural characteristics of the selected protein. The secondary parameters of the selected protein determined the alpha helix (blue color), extended strand (red color), beta-turn (light-green color), and random coil (light-yellow color).



(a)



(b)

Figure 3. (a) Amino acid types of the selected protein, and (b) Sequence plot of the selected protein.

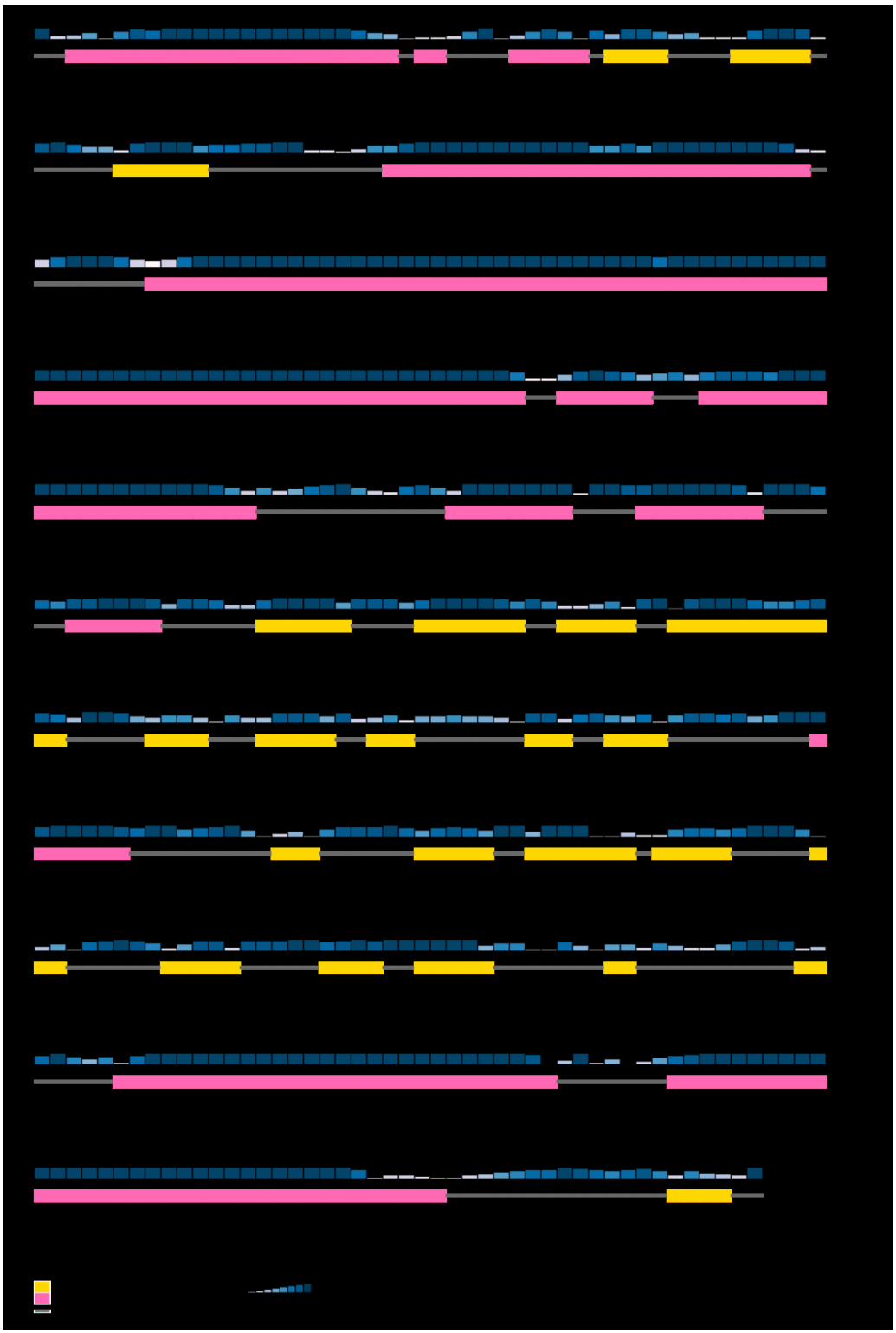
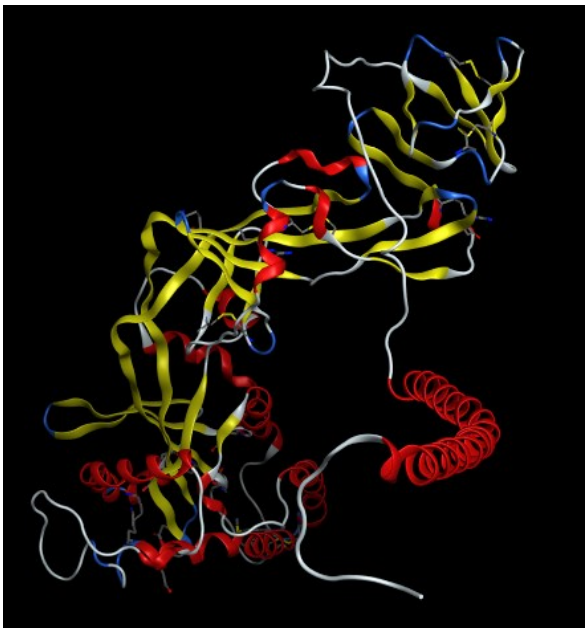


Figure 4. The secondary structure of the selected protein.

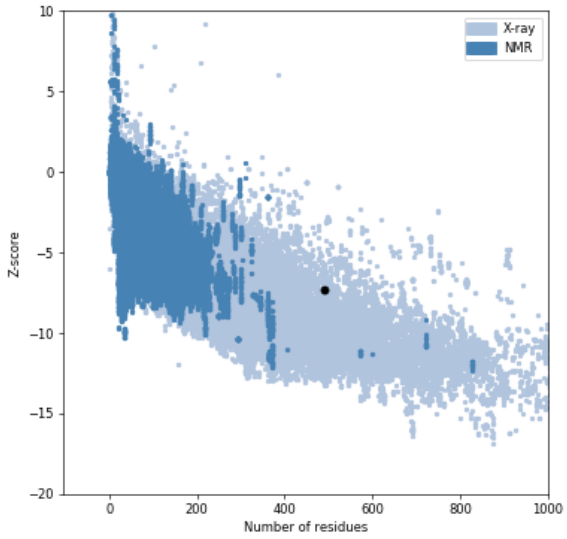
3.4 The three-dimensional protein structure anticipation and assessment

The three-dimensional form of a protein is known as its tertiary structure. One primary 'backbone' polypeptide chain in the tertiary structure comprises one or more protein secondary structures (PSSs) called domains [60-62]. There are a variety of possible interactions and bonds between amino acid side chains. The sequence-structure gap (SSG) is a significant obstacle in computational biology and chemistry, and protein structure anticipation is one strategy to close this gap. Accurately predicting the structure of a protein is critical since protein structure dictates its function [60, 63, 64]. The most favored protein templated (HHpred ID: 2B9B_A) was selected for anticipation of the three-dimensional protein structure by the Modeller program with the HHpred interface with the probability of 100%, E-value 2.8×10^{-132}, and target length of 497 (Figure 5).

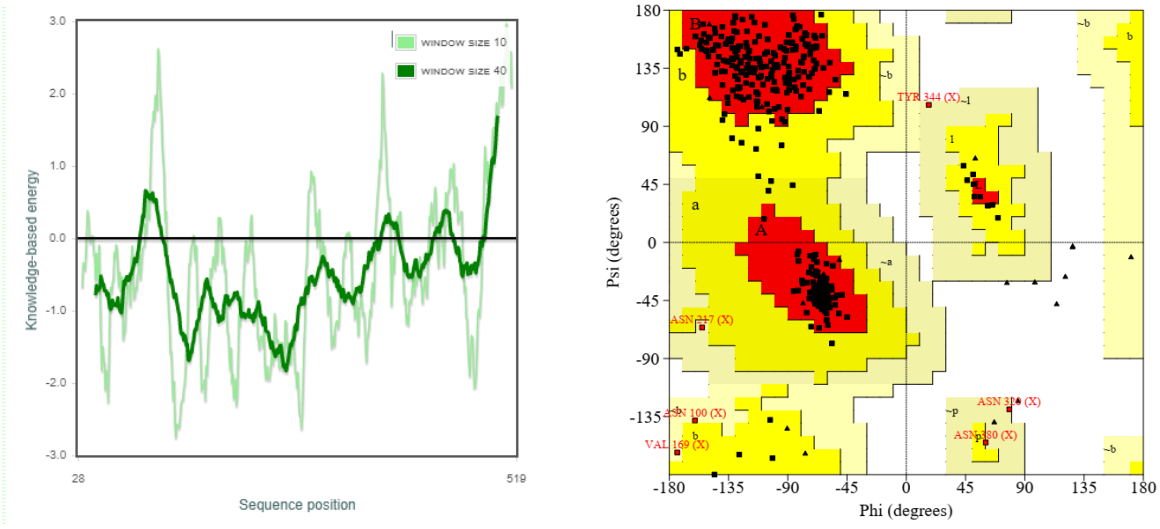
The estimated Ramachandran plot calculations of the selected protein were as residues in most favored regions (411, 91.9%), residues in additional allowed regions (30, 6.7%), residues in generously allowed regions (6, 1.3%), number of non-glycine and non-proline residues (447, 100.0%), and there was no residue in disallowed regions (Figure 5). Moreover, the local model assessment and the overall model quality by Z-score (-7.26) assessed the anticipated protein model quality and validated the structure of the protein.



(a)



(b)



(c) (d)

Figure 5. Tertiary structural anticipation and assessment of the selected protein. (a) The predicted three-dimensional structure, (b) The overall model quality by Z-score (-7.26), (c) The local model quality assessment, and (d) Ramachandran plot statistics obtained from the SAVES program.

5 Conclusions

NiV has developed as a fatal zoonotic disease. Bats, the natural reservoir of the virus, are adept at viral propagation and human outbreaks continue to be documented routinely. Since bats may be found worldwide, we might expect to see new epidemics in previously unaffected regions. Acute illness progression and a high death rate make a correct diagnosis challenging. The absence of accessible, affordable diagnostic tests and laboratories to process viral samples makes the situation worse. The total caseload is low, and the course of infection is rapid. Thus there is a dearth of investigations into human subjects that might yield effective therapy and prevention. The selected protein's secondary and tertiary characteristics demonstrated the protein structure-based relationships and, therefore, more comprehending properties of the protein. The protein is a fusion protein deeply associated with viral infection. Therefore, the selected protein can be a target for both protein-based drug and vaccine design against the protein to minimize viral infections.

References

1. Goh, K.J., et al., *Clinical features of Nipah virus encephalitis among pig farmers in Malaysia.* N Engl J Med, 2000. **342**(17): p. 1229-35.
2. Rahman, M. and A. Chakraborty, *Nipah virus outbreaks in Bangladesh: a deadly infectious disease.* WHO South East Asia J Public Health, 2012. **1**(2): p. 208-212.

3. Chadha, M.S., et al., *Nipah virus-associated encephalitis outbreak, Siliguri, India.* Emerg Infect Dis, 2006. **12**(2): p. 235-40.
4. Amarasinghe, G.K., et al., *Taxonomy of the order Mononegavirales: update 2019.* Arch Virol, 2019. **164**(7): p. 1967-1980.
5. Halpin, K., et al., *Pteropid bats are confirmed as the reservoir hosts of henipaviruses: a comprehensive experimental study of virus transmission.* Am J Trop Med Hyg, 2011. **85**(5): p. 946-51.
6. Epstein, J.H., et al., *Nipah virus: impact, origins, and causes of emergence.* Curr Infect Dis Rep, 2006. **8**(1): p. 59-65.
7. Banerjee, S., et al., *Nipah virus disease: A rare and intractable disease.* Intractable Rare Dis Res, 2019. **8**(1): p. 1-8.
8. Luby, S.P., et al., *Foodborne transmission of Nipah virus, Bangladesh.* Emerg Infect Dis, 2006. **12**(12): p. 1888-94.
9. Arunkumar, G., et al., *Outbreak Investigation of Nipah Virus Disease in Kerala, India, 2018.* J Infect Dis, 2019. **219**(12): p. 1867-1878.
10. Soman Pillai, V., G. Krishna, and M. Valiya Veettil, *Nipah Virus: Past Outbreaks and Future Containment.* Viruses, 2020. **12**(4).
11. Aguilar, H.C. and R.M. Iorio, *Henipavirus membrane fusion and viral entry.* Curr Top Microbiol Immunol, 2012. **359**: p. 79-94.
12. Ang, B.S.P., T.C.C. Lim, and L. Wang, *Nipah Virus Infection.* J Clin Microbiol, 2018. **56**(6).
13. Paul, L., *Nipah virus in Kerala: a deadly Zoonosis.* Clin Microbiol Infect, 2018. **24**(10): p. 1113-1114.
14. Hickey, A.C. and C.C. Broder, *The mechanism of henipavirus fusion: Examining the relationships between the attachment and fusion glycoproteins.* Virologica Sinica, 2009. **24**(2): p. 110-120.
15. Lee, B. and Z.A. Ataman, *Modes of paramyxovirus fusion: a Henipavirus perspective.* Trends Microbiol, 2011. **19**(8): p. 389-99.
16. Steffen, D.L., et al., *Henipavirus mediated membrane fusion, virus entry and targeted therapeutics.* Viruses, 2012. **4**(2): p. 280-308.
17. Yan, D. and Z. Wang, *[Research Progress in Enveloped Glycoproteins and the Membrane-fusion Mechanism of Nipah Virus].* Bing Du Xue Bao, 2016. **32**(3): p. 361-8.
18. Mathieu, C. and B. Horvat, *Henipavirus pathogenesis and antiviral approaches.* Expert Rev Anti Infect Ther, 2015. **13**(3): p. 343-54.
19. Sayers, E.W., et al., *Database resources of the National Center for Biotechnology Information.* Nucleic Acids Res, 2021. **49**(D1): p. D10-d17.
20. Schneider, M., M. Tognolli, and A. Bairoch, *The Swiss-Prot protein knowledgebase and ExPASy: providing the plant community with high quality proteomic data and tools.* Plant Physiol Biochem, 2004. **42**(12): p. 1013-21.
21. Stothard, P., *The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences.* Biotechniques, 2000. **28**(6): p. 1102, 1104.
22. Geourjon, C. and G. Deléage, *SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments.* Bioinformatics, 1995. **11**(6): p. 681-684.

23. Moffat, L. and D.T. Jones, *Increasing the accuracy of single sequence prediction methods using a deep semi-supervised learning framework.* Bioinformatics, 2021. **37**(21): p. 3744-3751.
24. Webb, B. and A. Sali, *Comparative Protein Structure Modeling Using MODELLER.* Curr Protoc Bioinformatics, 2016. **54**: p. 5.6.1-5.6.37.
25. Levine, T.P., et al., *Discovery of new Longin and Roadblock domains that form platforms for small GTPases in Ragulator and TRAPP-II.* Small GTPases, 2013. **4**(2): p. 62-9.
26. Zimmermann, L., et al., *A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core.* Journal of Molecular Biology, 2018. **430**(15): p. 2237-2243.
27. Laskowski, R.A., et al., *AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR.* J Biomol NMR, 1996. **8**(4): p. 477-86.
28. Wiederstein, M. and M.J. Sippl, *ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins.* Nucleic Acids Research, 2007. **35**(suppl_2): p. W407-W410.
29. Nham Tran, T.L., et al., *Physicochemical Characteristics of Protein Isolated from Thraustochytrid Oilcake.* Foods, 2020. **9**(6).
30. Dey, D., et al., *Natural flavonoids effectively block the CD81 receptor of hepatocytes and inhibit HCV infection: a computational drug development approach.* Molecular Diversity, 2022.
31. Lu, Z.X., et al., *Composition, physicochemical properties of pea protein and its application in functional foods.* Crit Rev Food Sci Nutr, 2020. **60**(15): p. 2593-2605.
32. Karadag, M., et al., *Physicochemical determinants of antibody-protein interactions.* Adv Protein Chem Struct Biol, 2020. **121**: p. 85-114.
33. Khan, R.A., et al., *Diterpenes/Diterpenoids and Their Derivatives as Potential Bioactive Leads against Dengue Virus: A Computational and Network Pharmacology Study.* Molecules, 2021. **26**(22).
34. Kontermann, R.E., *Half-life extended biotherapeutics.* Expert Opin Biol Ther, 2016. **16**(7): p. 903-15.
35. Sleep, D., *Albumin and its application in drug delivery.* Expert Opin Drug Deliv, 2015. **12**(5): p. 793-812.
36. Kontermann, R.E., *Strategies for extended serum half-life of protein therapeutics.* Curr Opin Biotechnol, 2011. **22**(6): p. 868-76.
37. Saikat, A.S.M., *An In Silico Approach for Potential Natural Compounds as Inhibitors of Protein CDK1/Cks2.* Chemistry Proceedings, 2022. **8**(1): p. 5.
38. Zaman, R., et al., *Current strategies in extending half-lives of therapeutic proteins.* J Control Release, 2019. **301**: p. 176-189.
39. Louw, A., *GR Dimerization and the Impact of GR Dimerization on GR Protein Stability and Half-Life.* Front Immunol, 2019. **10**: p. 1693.
40. Werle, M. and A. Bernkop-Schnürch, *Strategies to improve plasma half life time of peptide and protein drugs.* Amino Acids, 2006. **30**(4): p. 351-67.
41. Podust, V.N., et al., *Extension of in vivo half-life of biologically active molecules by XTEN protein polymers.* J Control Release, 2016. **240**: p. 52-66.
42. Rajib, H., et al., *Amentoflavone, New Hope against SARS-CoV-2: An Outlook through its Scientific Records and an in silico Study.* Pharmacognosy Research, 2021. **13**(3).

43. Niu, X., et al., *Interconnection between the protein solubility and amino acid and dipeptide compositions.* Protein Pept Lett, 2013. **20**(1): p. 88-95.
44. Huang, H.L., et al., *Prediction and analysis of protein solubility using a novel scoring card method with dipeptide composition.* BMC Bioinformatics, 2012. **13 Suppl 17**(Suppl 17): p. S3.
45. Saikat, A.S.M., et al., *Structural and Functional Elucidation of IF-3 Protein of Chloroflexus aurantiacus Involved in Protein Biosynthesis: An In Silico Approach.* BioMed Research International, 2021. **2021**: p. 9050026.
46. Yagasaki, M. and S. Hashimoto, *Synthesis and application of dipeptides; current status and perspectives.* Appl Microbiol Biotechnol, 2008. **81**(1): p. 13-22.
47. Guruprasad, K., B.V. Reddy, and M.W. Pandit, *Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence.* Protein Eng, 1990. **4**(2): p. 155-61.
48. Saikat, A.S.M., et al., *Structural and Functional Annotation of Uncharacterized Protein NCGM946K2_146 of Mycobacterium Tuberculosis: An In-Silico Approach.* Proceedings, 2020. **66**(1): p. 13.
49. Gamage, D.G., et al., *Applicability of Instability Index for In vitro Protein Stability Prediction.* Protein Pept Lett, 2019. **26**(5): p. 339-347.
50. Panda, S. and G. Chandra, *Physicochemical characterization and functional analysis of some snake venom toxin proteins and related non-toxin proteins of other chordates.* Bioinformation, 2012. **8**(18): p. 891-6.
51. Ikai, A., *Thermostability and aliphatic index of globular proteins.* J Biochem, 1980. **88**(6): p. 1895-8.
52. Enany, S., *Structural and functional analysis of hypothetical and conserved proteins of Clostridium tetani.* Journal of Infection and Public Health, 2014. **7**(4): p. 296-307.
53. Chang, K.Y. and J.R. Yang, *Analysis and prediction of highly effective antiviral peptides based on random forests.* PLoS One, 2013. **8**(8): p. e70166.
54. Kyte, J. and R.F. Doolittle, *A simple method for displaying the hydropathic character of a protein.* J Mol Biol, 1982. **157**(1): p. 105-32.
55. Thomas, S., et al., *CAMP: a useful resource for research on antimicrobial peptides.* Nucleic Acids Res, 2010. **38**(Database issue): p. D774-80.
56. Hossain, R., et al., *Biosynthesis of Secondary Metabolites Based on the Regulation of MicroRNAs.* Biomed Res Int, 2022. **2022**: p. 9349897.
57. Gromiha, M.M., *Chapter 1 - Proteins*, in *Protein Bioinformatics*, M.M. Gromiha, Editor. 2010, Academic Press: Singapore. p. 1-27.
58. Wardah, W., et al., *Protein secondary structure prediction using neural networks and deep learning: A review.* Comput Biol Chem, 2019. **81**: p. 1-8.
59. Koch, I. and T. Schäfer, *Protein super-secondary structure and quaternary structure topology: theoretical description and application.* Curr Opin Struct Biol, 2018. **50**: p. 134-143.
60. Jisna, V.A. and P.B. Jayaraj, *Protein Structure Prediction: Conventional and Deep Learning Perspectives.* Protein J, 2021. **40**(4): p. 522-544.
61. Hou, J., et al., *Protein tertiary structure modeling driven by deep learning and contact distance prediction in CASP13.* Proteins, 2019. **87**(12): p. 1165-1178.
62. Tamburrini, K.C., et al., *Predicting Protein Conformational Disorder and Disordered Binding Sites.* Methods Mol Biol, 2022. **2449**: p. 95-147.

63. Reinert, Z.E. and W.S. Horne, *Protein backbone engineering as a strategy to advance foldamers toward the frontier of protein-like tertiary structure.* Org Biomol Chem, 2014. **12**(44): p. 8796-802.
64. Shimizu, K., et al., *Comparative analysis of membrane protein structure databases.* Biochim Biophys Acta Biomembr, 2018. **1860**(5): p. 1077-1091.