chemistry proceedings

MDPI

Proceeding Paper

Computational Approaches for Molecular Characterization and Structure-Based Functional Elucidation of the Matrix Protein from *Nipah henipavirus* †

Abu Saim Mohammad Saikat1 [1,*], Apurbo Kumar Paul [2] and Dipta Dey [1]

[1] Department of Biochemistry and Molecular Biology, Life Science Faculty, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj 8100, Bangladesh; diptadey727@gmail.com
[2] Department of Materials Science and Engineering, University of Rajshahi, Rajshahi 6205, Bangladesh; apurbomse@protonmail.com
* Correspondence: asmsaikat.bmb@gmail.com; Tel.: +880-1795840353
† Presented at the 26th International Electronic Conference on Synthetic Organic Chemistry; Available online: https://ecsoc-26.sciforum.net.

Abstract: *Nipah henipavirus* is an emerging RNA virus that poses a danger to world safety due to its high fatality rate. The Nipah virus has caused several illness epidemics in South and Southeast Asia. The matrix protein of *Nipah henipavirus* plays a crucial function in linking the viral envelope to the virus core. Connecting the viral envelope to the virus core is critical for virus assembly. Through analyses of structural and functional protein explanations, bioinformatics tools can aid in our comprehension of the protein. This study intends to provide structural and functional annotations to proteins. Using in-silico approaches, the analysis also assigns the protein's physicochemical properties, three-dimensional structure, and functional annotation. The in-silico research validated the protein's hydrophilic nature and alpha (α) helix-dominated secondary structure. The protein's tertiary structure model is generally consistent based on various quality evaluation approaches. The functional explanation claimed that the protein is a structural protein that connects the viral envelope to the virus core, a protein that is necessary for virus assembly. This study reveals the importance of the matrix protein as a functional protein needed by *Nipah henipavirus*.

Keywords: *Nipah henipavirus*; matrix protein; protein nature; functional annotation; protein characterization

Citation: Saikat, A.S.M.; Paul, A.K.; Dey, D. Computational Approaches for Molecular Characterization and Structure-based Functional Elucidation of the Matrix Protein from *Nipah henipavirus*. **2022**, *4*, x. https://doi.org/10.3390/xxxxx

Academic Editor(s):

Published: 15 November 2022

1. Introduction

Nipah henipavirus is a bat-borne virus that can infect humans and other animals [1,2]. Nipah virus infection is a zoonotic disease that is spread from animals to humans. It can also be spread through contaminated food or from person to person. It causes a variety of symptoms in infected individuals, ranging from asymptomatic (subclinical) infection to acute respiratory sickness and deadly encephalitis [3–5]. Nipah virus is a member of the Paramyxoviridae family and the genus Henipavirus along with the Hendra virus, which has also caused disease outbreaks [6,7].The Nipah virus genome is a single (nonsegmented) negative-sense, single-stranded RNA with a length of over 18 kb, far longer than that of other paramyxoviruses [8,9]. The Nipah virus was initially detected in pigs and pig farmers in Peninsular Malaysia in 1998 [9]. Infection outbreaks of the Nipah virus have been documented in Malaysia, Singapore, Bangladesh, and India [6]. The highest rates of death from Nipah virus infection have been reported in Bangladesh, where outbreaks are most common in the winter [10,11]. Consumption of fruits or fruit products (such as raw date palm juice) contaminated with urine or saliva from infected fruit bats was the most likely cause of infection in later outbreaks in Bangladesh and India [11].

About 700 human cases of Nipah virus had been reported as of May 2018, with 50 to 75 percent of those affected dying [12]. In the Indian state of Kerala, an epidemic of the disease led to 17 deaths in May 2018 [13].

A study of the proteins using bioinformatics tools allows researchers to assess their three-dimensional structural conformation, classify new domains, explore certain pathways to gain a better understanding of our evolutionary tree, uncover more clusters, and assign roles to the proteins [14–16]. This knowledge can also be used to develop successful pharmacological methods and aid in the development of new drugs to treat a wide range of diseases [17–19]. This study demonstrated the matrix protein secondary as well as tertiary characteristics that are associated with protein-structure relationships. The selected protein can be used as a potential target for protein-based drug and protein-based vaccine design candidates to minimize the viral infection.

2. Materials and Methods

2.1. Protein Selection and Sequence Retrieval

The amino acid (aa) sequence of the matrix protein found in *Nipah henipavirus* was obtained in FASTA format from the NCBI database [20].

2.2. Physicochemical Characterization of the Selected Protein

The amino acid sequence composition, instability index, aliphatic index, GRAVY (assessment of hydrophobicity or hydrophilicity of a protein), and extinction coefficients were all measured using the ExPASy server ProtParam tool [21]. The theoretical isoelectric point (pI) of the QBQ56721.1 protein was also measured using SMS Suite (v.2.0) [22].

2.3. Functional Annotation of the Selected Protein

The conserved domain in the protein QBQ56721.1 was predicted using the NCBI platform's CD-search tool [23,24]. The ExPASy software's ScanProsite tool [25], Pfam tool [26] were used to determine protein motifs. The evolutionary relationships of the protein QBQ56721.1 were assigned by the SuperFamily program [27].

2.4. Secondary Structural Properties and Assessment

The self-optimized prediction method with alignment (SOPMA) is used to predict secondary structure elements [28,29]. The secondary structure was predicted using the SPIPRED (v.4.0) [30] algorithm.

2.5. Three-Dimensional Structure Prediction and Validation of the Selected Protein

With Modeller [31], HHpred [32] predicted the three-dimensional (tertiary) structure. The most suitable template (HHpred ID: 6BK6 A) was chosen for creating the tertiary structure, with probability, E value, aligned cols, and goal lengths of 100, 2.4 × 10-116, 342, and 372 correspondingly. To predict the Ramachandran plot and validate the expected tertiary structure, the PROCHECK tool of the SAVES (v.6.0 program) [33,34] was used.

3. Results and Discussion

3.1. Protein Sequence Retrieval

The amino acid (aa) sequence of *Nipah henipavirus* protein (QBQ56721.1) was obtained from the NCBI database. The 352-amino-acid-long protein sequence was used to model the tertiary structure of the protein QBQ56721.1. Table 1 provides additional information on the protein (QBQ56721.1).

Table 1. Protein retrieval.

Protein Individualities	Protein Information
Locus	QBQ56721
Amino acid	352 aa
Definition	matrix protein [*Nipah henipavirus*]
Accession	QBQ56721
Version	QBQ56721.1
Source	*Nipah henipavirus*
Organism	*Nipah henipavirus*
FASTA sequence	>QBQ56721.1 matrix protein [*Nipah henipavirus*] MEPDIKSISSESMEGVSDFSPSSWENGGYLDKVEPEIDENGSMIPKYKIYTPGANERKYNN YMYLICYGF VEDVERTPETGKRKKIRTIAAYPLGVGKSASHPQDLLEELCSLKVTVRRTAGSTEKVVFG SSGPLNHLVP WKKVLTGGSIFNAVKVCRNVDQIQLDKHQALRIFFLSITKLNDSGIYMIPRTMLEFRRNN AIAFNLLVYL KIDADLSKMGIQGSLDKDGFKVASFMLHLGNFVRRAGKYYSVDYCRRKIDRMKLQFSL GSIGGLSLHIKI NGVISKRLFAQMGFQKNLCFSLMDINPWLNRLTWNNSCEISRVAAVLQPSVPREFMIYD DVFIDNTGRIL KG

3.2. Physicochemical Properties Identification of the Protein

The amino acid sequence of QBQ56721.1, which is found in *Nipah henipavirus*, was obtained in FASTA format and utilized as a query sequence for physicochemical parameter measurement. The protein is stable because its instability index is 30.59 (less than 40.00) [35]. The theoretical isoelectric point (pI) of the protein (pI 9.31, 9.65*) indicates that it is basic [36–38]. The molecular weight, aliphatic index, instability index, and GRAVY are 39,847.16 Dalton, 89.69, 30.59 and −0.212 respectively (Table 2) [39–41]. The protein's higher aliphatic index value of 89.69 indicates increased thermos-stability over a wide temperature range, which is a favorable factor [42,43]. The GRAVY index value of -0.212 suggested the protein's hydrophilic character and hence the prospect of more water interaction [44,45].

Table 3. Physicochemical parameters.

Parameters	Value
Molecular weight	39,847.16
Theoretical pI	9.31, 9.65 *
Total number of negatively charged residues (Asp+Glu)	36
Total number of positively charged residues (Arg+Lys)	48
Formula	$C_{1787}H_{2831}N_{485}O_{510}S_{18}$
Total number of atoms	5631
The estimated half-life	(a) 30 h (mammalian reticulocytes, in vitro). (b) >20 h (yeast, in vivo). (c) >10 h (Escherichia coli, in vivo).
Instability index (II)	30.59
Aliphatic index	89.69
Grand average of hydropathicity (GRAVY)	−0.212

* pI calculated by the SMS v.2.0.

3.3. Functional Annotation Anticipation of the Selected Protein

The NCBI CDD tool identifies the domain that appears in identical protein sequences. RPS-BLAST is used by CD-Search to compare a test sequence to position-specific rating datasets compiled from conserved domain (CD) alignments in the CD protein cluster. The CD search engine identified a conserved domain in the protein QBQ56721.1 as a viral matrix protein (matrix, accession no. pfam00661). Viral matrix proteins are structural proteins that connect the viral envelope and the virus core [46,47]. The matrix protein plays an important role in virus assembly and linking the viral envelope with the virus core. It's possible that they are found in Morbillivirus, Paramyxovirus, and Pneumovirus [47,48]. A motif was also predicted by the Pfam software at locations 16–349 (Pfam ID: PF00661; Viral matrix protein; e value of 1:7 10146). Protein motifs are small regions of three-dimensional protein structure or amino acid sequence that are shared by multiple proteins [48,49]. Motifs are distinct regions of protein structure that may or may not be defined by a distinct chemical or biological function [48,50–53].

The CDD technique also confirmed the presence of viral matrix protein at the 17–348 position. The lone member of the superfamily cl02918 is the viral matrix protein (CDD no. pfam00661). A protein superfamily is a group of proteins made up of one or more protein families [46,53,54]. The set of all superfamilies must be a partitioning of the set of all protein sequences or subsequences defined by the protein families relationship, and each superfamily must be closed under transitivity [54].The protein QBQ56721.1 (Figure 1) was predicted to be closely related to the matrix superfamily by the SuperFamily tool (e value of 0.0).Main text paragraph (M_Text).

Figure 1. Functional annotation of the selected protein QBQ56721.1. The graphical summary represents the conserved domains identified inthe query sequence. The aligned sequences represent the conserved domains identified on the query sequence by comparing with theconserved protein

domain family, matrix (CDD accession no. pfam00661). The Pfam software predicted a motif at 16–349 (accession no. PF00661) as viral matrix protein. The SuperFamily program predicted the protein as a member of the matrix superfamily.

3.4. Secondary Structural Inquiry

Protein structure and function is inextricably linked. Helix, coil, sheet, and turn are secondary structural elements that have a fantastic association with protein function, structure, and engagement [55–58]. The secondary-structural element of the protein (QBQ56721.1) was predicted by the SOPMA software when the alpha helix (Hh), extended strand (Ee), beta turn (Tt), and random coil (Cc) were 60 (17.05%), 87 (24.72%), 20 (5.68%), and 185 (52.56%), respectively (Table 5). The PSIPRED software predicts the helix, strand, and coil of the matrix protein (QBQ56721.1) with more confidence (Figure 1). Table 4 shows the amino acid composition obtained from the ExPASy service's ProtParam Tool.

Table 4. Amino acid composition.

Amino Acids	Percentage (%)
Ala (A)	5.2%
Arg (R)	2.1%
Asn (N)	9.4%
Asp (D)	5.7%
Cys (C)	0.2%
Gln (Q)	5.2%
Glu (E)	8.5%
Gly (G)	4.8%
His (H)	1.3%
Ile (I)	6.3%
Leu (L)	9.2%
Lys (K)	9.1%
Met (M)	2.0%
Phe (F)	4.0%
Pro (P)	3.0%
Ser (S)	6.9%
Thr (T)	6.9%
Trp (W)	0.6%
Tyr (Y)	3.0%
Val (V)	6.7%

Table 5. Secondary structural elements.

Secondary Structure Elements	Values (%)
Alpha helix (Hh)	60 (17.05)
3_{10} helix (Gg)	0
Pi helix (Ii)	0
Beta bridge (Bb)	0
Extended strand (Ee)	87 (24.72)
Beta turn (Tt)	20 (5.64)
Bend region (Ss)	0
Random coil (Cc)	185 (52.56)
Ambiguous states	0
Other states	0

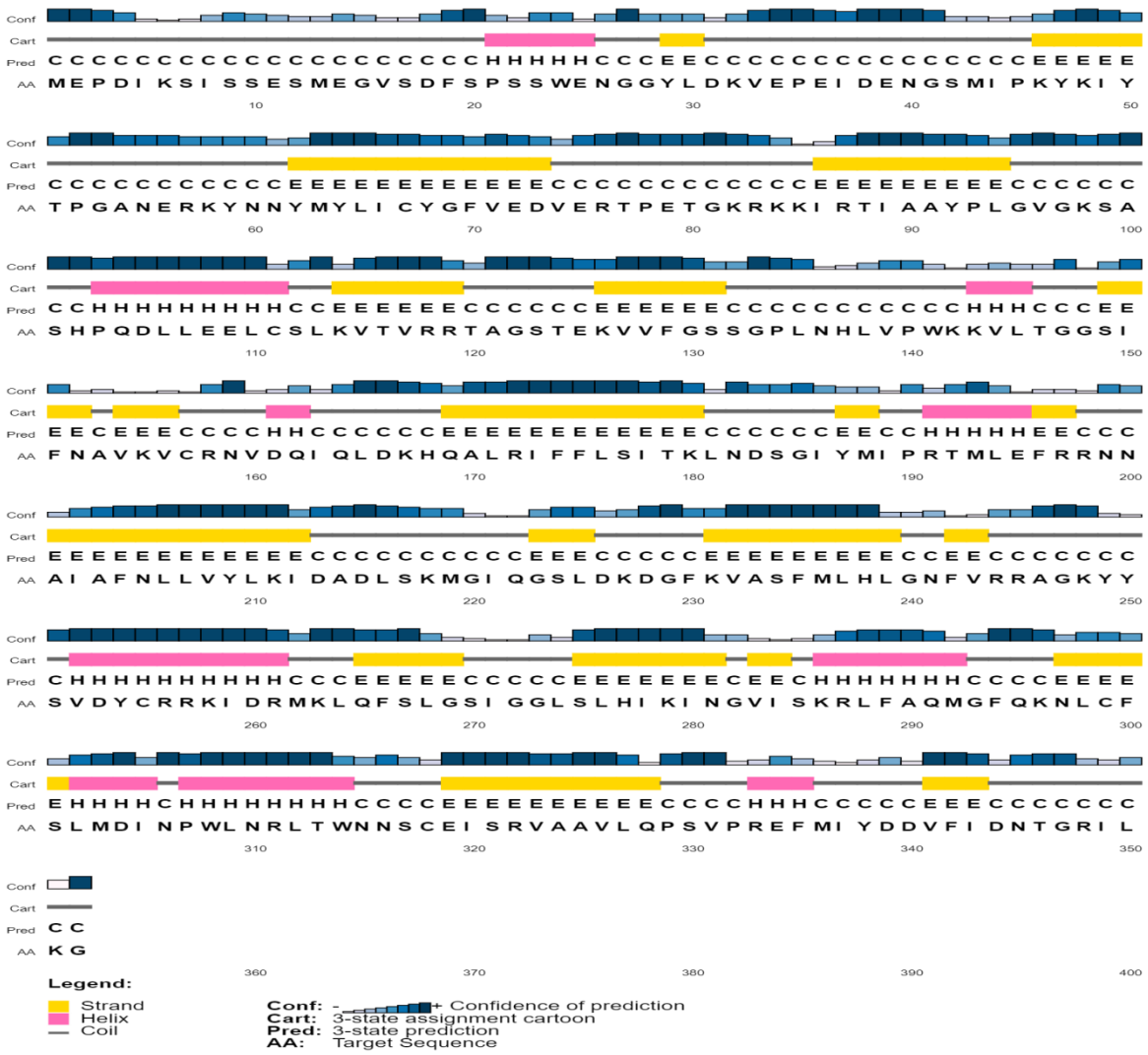

Figure 2. Predicted secondary structure of the selected protein.

3.5. Tertiary-Structure Anticipation and Validation of the Protein

The target sequence of QBQ56721.1 in FASTA format was inserted into the HHpred Template Selection tool as input, and the most suitable template (6BK6 A) was selected with a probability rate of 100%, E-Value of 2.4 × 10 -116, Cols of 342, and the target length of 372, and finally stored the tertiary modeled protein structure in PDB format predicted by Modeller (Figure 3). The Ramachandran plot by PROCHECK (Figure 4) was used to assess the matrix protein's tertiary structure, which revealed that 92.4 percent of the total residues (342) were found in the core [A,B,L]; 6.3 percent of residues were in the additional allowed regions [a,b,l,p]; and 0.7 percent of residues were in the generously allowed regions [a,b,l,p]. The total number of non-glycine and non-proline residues was 301; the end-residues (excluding Gly and Pro) were 1.0; the glycine and proline residues were 27 and 13, respectively, out of 473 total residues (Table 7). Verify 3D; a tertiary structure evaluation tool was used to demonstrate that the anticipated tertiary structure passed the evaluation.

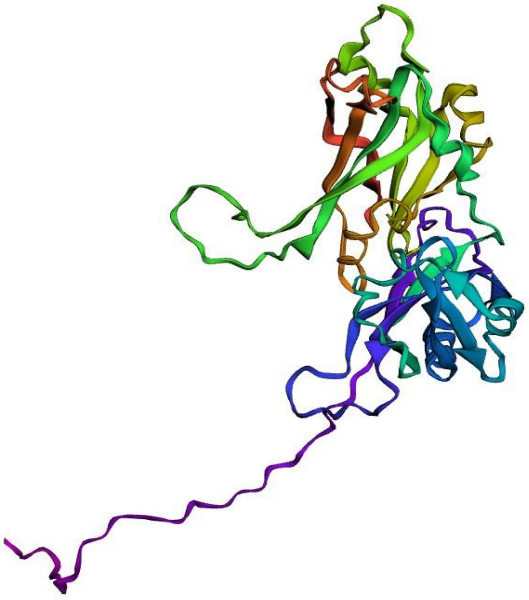

Figure 3. Predicted tertiary structure by HHpred tool employing the Modeller application.

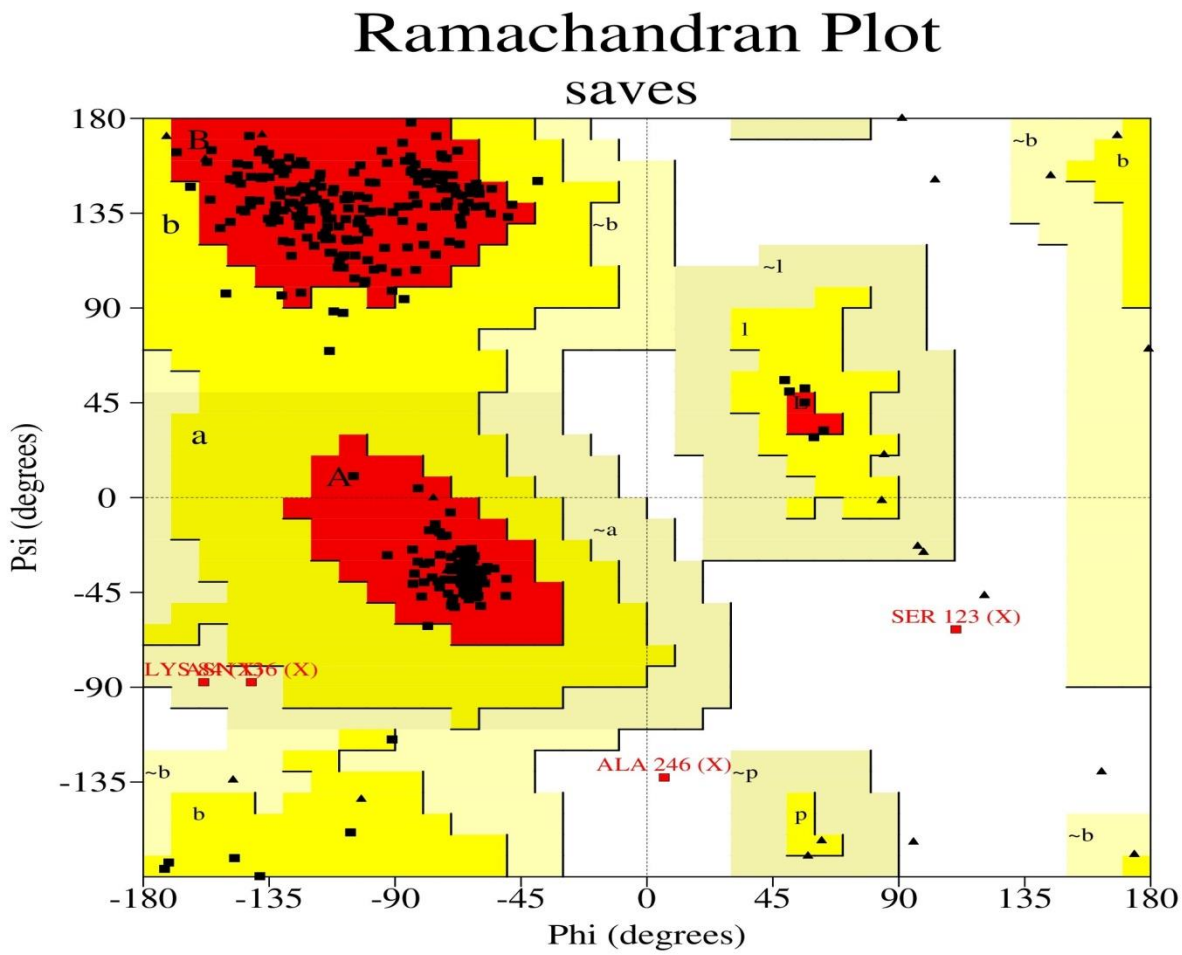

Figure 4. The Ramachandran plot statistics of Modeller-predicted three-dimensional protein structure validated by the PROCHECK program.

Table 4. Ramachandran plot statistics of the modeled protein.

Ramachandran Plot Statistics	Value (%)
Residues in the most favored regions [A, B, L]	278 (92.4)
Residues in additional allowed regions [a, b, l, p]	19(6.3)
Residues in generously allowed regions [~ a, ~ b, ~ l, ~ p]	2 (0.7)
Residues in disallowed regions	2 (0.7)
Number of nonglycine and nonproline residues	301
Number of end residues (excl. Gly and Pro)	1
Number of glycine residues (shown as triangles)	27
Number of proline residues	13
Total number of residues	342

4. Conclusions

Understanding how proteins function is vital for describing how they work, and this protein is critical for virus assembly. With the virus core, matrix protein binds to the viral envelope. This research reveals the protein's fundamental features, such as its hydrophilic nature and functional annotation, in relation to its tertiary structure. As a result, the outcomes of this study demonstrate the efficacy and scope of future research on the matrix protein using the bioinformatics methodologies used in this investigation. The selected protein secondary and tertiary structures demonstrated the protein-function relationships of the matrix protein. This research will strengthen and sharpen our understanding of pathophysiology, allowing for the development of protein-based promising drugs and vaccines candidates to combat Nipah virus infection.

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/xxx/s1, Figure S1: title; Table S1: title; Video S1: title.

Author Contributions: Conceptualization, A.S.M.S.; methodology, A.S.M.S. and A.K.P.; software, A.S.M.S.; validation, A.S.M.S., D.D.; formal analysis, A.S.M.S.; investigation, A.S.M.S.; resources, A.S.M.S., A.K.P.; data curation, A.S.M.S., A.K.P., and D.D.; writing—original draft preparation, A.S.M.S., A.K.P, and D.D.; writing—review and editing, A.S.M.S.; visualization, A.S.M.S., A.K.P, and D.D.; supervision, A.S.M.S.; project administration, A.S.M.S.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ang, B.S.P.; Lim, T.C.C.; Wang, L. Nipah Virus Infection. *J. Clin. Microbiol.* **2018**, *56*.
2. Paul, L. Nipah virus in Kerala: A deadly Zoonosis. *Clin. Microbiol. Infect.* **2018**, *24*, 1113–1114.
3. Aditi; Shariff, M. Nipah virus infection: A review. *Epidemiol. Infect.* **2019**, *147*, e95.
4. Sharma, V.; et al. Emerging trends of Nipah virus: A review. *Rev. Med. Virol.* **2019**, *29*, e2010.
5. Soman Pillai, V.; Krishna, G.; Veettil, M.V. Nipah Virus: Past Outbreaks and Future Containment. *Viruses* **2020**, *12*, 465.
6. Lo, M.K.; Rota, P.A. The emergence of Nipah virus, a highly pathogenic paramyxovirus. *J. Clin. Virol.* **2008**, *43*, 396–400.
7. Ternhag, A.; Penttinen, P. Nipah virus--another product from the Asian "virus factory". *Lakartidningen* **2005**, *102*, 1046–1047.
8. Choi, C. Nipah's return. The lethal "flying fox" virus may spread between people. *Sci. Am.* **2004**, *291*, 21a, 22.
9. Singh, R.K.; et al. Nipah virus: Epidemiology, pathology, immunobiology and advances in diagnosis, vaccine designing and control strategies—A comprehensive review. *Vet. Q.* **2019**, *39*, 26–55.
10. Epstein, J.H.; et al. Nipah virus dynamics in bats and implications for spillover to humans. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 29190–29201.

11. Yadav, P.D.; et al. Nipah Virus Sequences from Humans and Bats during Nipah Outbreak, Kerala, India, 2018. *Emerg. Infect. Dis.* **2019**, *25*, 1003–1006.
12. Sudeep, A.B.; et al. Detection of Nipah virus in Pteropus medius in 2019 outbreak from Ernakulam district, Kerala, India. *BMC Infect. Dis.* **2021**, *21*, 162.
13. Yadav, P.D.; et al. Detection of Nipah virus RNA in fruit bat (Pteropus giganteus) from India. *Am. J. Trop. Med. Hyg.* **2012**, *87*, 576–578.
14. Gaudino, M.; et al. High Pathogenicity of Nipah Virus from Pteropus lylei Fruit Bats, Cambodia. *Emerg. Infect. Dis.* **2020**, *26*, 104–113.
15. Rathish, B.; Vaishnani, K. Nipah Virus. In *StatPearls*; StatPearls Publishing Copyright © 2022; StatPearls Publishing LLC.: Treasure Island, FL, USA, 2022.
16. Looi, L.M.; Chua, K.B. Lessons from the Nipah virus outbreak in Malaysia. *Malays. J. Pathol.* **2007**, *29*, 63–67.
17. Lam, S.K.; Chua, K.B. Nipah virus encephalitis outbreak in Malaysia. *Clin. Infect. Dis.* **2002**, *34* (Suppl. 2), S48–S51.
18. Singhai, M.; et al. Nipah Virus Disease: Recent Perspective and One Health Approach. *Ann. Glob. Health* **2021**, *87*, 102.
19. Gómez Román, R.; et al. Nipah@20: Lessons Learned from Another Virus with Pandemic Potential. *mSphere* **2020**, *5*.
20. Sayers, E.W.; et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2021**, *49*, D10–D17.
21. Gasteiger, E.; et al. Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*; Walker, J.M., Ed.; Humana Press: Totowa, NJ, USA, 2005; pp. 571–607.
22. Stothard, P. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* **2000**, *28*, 1102–1104.
23. Lu, S.; et al. CDD/SPARCLE: The conserved domain database in 2020. *Nucleic Acids Res.* **2020**, *48*, D265–D268.
24. Marchler-Bauer, A.; et al. CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* **2017**, *45*, D200–D203.
25. Sigrist, C.J.; et al. New and continuing developments at PROSITE. *Nucleic Acids Res.* **2013**, *41*, D344–D347.
26. Sigrist, C.J.; et al. PROSITE: A documented database using patterns and profiles as motif descriptors. *Brief. Bioinform.* **2002**, *3*, 265–274.
27. Wilson, D.; et al. The SUPERFAMILY database in 2007: Families and functions. *Nucleic Acids Res.* **2007**, *35*, D308–D313.
28. Geourjon, C.; Deléage, G. SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput. Appl. Biosci.* **1995**, *11*, 681–684.
29. Deléage, G. ALIGNSEC: Viewing protein secondary structure predictions within large multiple sequence alignments. *Bioinformatics* **2017**, *33*, 3991–3992.
30. Moffat, L.; Jones, D.T. Increasing the accuracy of single sequence prediction methods using a deep semi-supervised learning framework. *Bioinformatics* **2021**, *37*, 3744–3751.
31. Webb, B.; Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinform.* **2016**, *54*, 5.6.1–5.6.37.
32. Gabler, F.; et al. Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. *Curr. Protoc. Bioinform.* **2020**, *72*, e108.
33. 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*, 477–486.
34. Bowie, J.U.; Lüthy, R.; Eisenberg, D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science* **1991**, *253*, 164–170.
35. Gamage, D.G.; et al. Applicability of Instability Index for In vitro Protein Stability Prediction. *Protein Pept. Lett.* **2019**, *26*, 339–347.
36. Pihlasalo, S.; et al. Method for estimation of protein isoelectric point. *Anal. Chem.* **2012**, *84*, 8253–8258.
37. Audain, E.; et al. Accurate estimation of isoelectric point of protein and peptide based on amino acid sequences. *Bioinformatics* **2016**, *32*, 821–827.
38. Saikat, A.S.M. An In Silico Approach for Potential Natural Compounds as Inhibitors of Protein CDK1/Cks2. *Chem. Proc.* **2022**, *8*, 5.
39. Wilkins, M.R.; Williams, K.L. Cross-species protein identification using amino acid composition, peptide mass fingerprinting, isoelectric point and molecular mass: A theoretical evaluation. *J. Theor. Biol.* **1997**, *186*, 7–15.
40. Kyte, J.; Doolittle, R.F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, *157*, 105–132.
41. 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*, 6821.
42. Ikai, A. Thermostability and aliphatic index of globular proteins. *J. Biochem.* **1980**, *88*, 1895–1898.
43. Dey, D.; et al. Natural flavonoids effectively block the CD81 receptor of hepatocytes and inhibit HCV infection: A computational drug development approach. *Mol. Divers.* **2022**.
44. Jin, Y.-T.; et al. Quantitative elucidation of associations between nucleotide identity and physicochemical properties of amino acids and the functional insight. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 4042–4048.
45. 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*, 13.
46. 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 Res. Int.* **2021**, *2021*, 9050026.

47. Battisti, A.J.; et al. Structure and assembly of a paramyxovirus matrix protein. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 13996–14000.
48. Shtykova, E.V.; et al. Solution Structure, Self-Assembly, and Membrane Interactions of the Matrix Protein from Newcastle Disease Virus at Neutral and Acidic pH. *J. Virol.* **2019**, *93*, e01450-18.
49. Stollar, E.J.; Smith, D.P. Uncovering protein structure. *Essays Biochem.* **2020**, *64*, 649–680.
50. Heizinger, L.; Merkl, R. Evidence for the preferential reuse of sub-domain motifs in primordial protein folds. *Proteins* **2021**, *89*, 1167–1179.
51. Xie, J.; Lai, L. Protein topology and allostery. *Curr. Opin. Struct. Biol.* **2020**, *62*, 158–165.
52. Santhouse, J.R.; Rao, S.R.; Horne, W.S. Analysis of folded structure and folding thermodynamics in heterogeneous-backbone proteomimetics. *Methods Enzymol.* **2021**, *656*, 93–122.
53. Vishwanath, S.; de Brevern, A.G.; Srinivasan, N. Same but not alike: Structure, flexibility and energetics of domains in multi-domain proteins are influenced by the presence of other domains. *PLoS Comput. Biol.* **2018**, *14*, e1006008.
54. Berezovsky, I.N.; Guarnera, E.; Zheng, Z. Basic units of protein structure, folding, and function. *Prog. Biophys. Mol. Biol.* **2017**, *128*, 85–99.
55. Padjasek, M.; et al. Structural zinc binding sites shaped for greater works: Structure-function relations in classical zinc finger, hook and clasp domains. *J. Inorg. Biochem.* **2020**, *204*, 110955.
56. Zhang, G.J.; et al. Secondary Structure and Contact Guided Differential Evolution for Protein Structure Prediction. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2020**, *17*, 1068–1081.
57. Rademaker, D.; et al. The Future of Protein Secondary Structure Prediction Was Invented by Oleg Ptitsyn. *Biomolecules* **2020**, *10*, 910.
58. Wardah, W.; et al. Protein secondary structure prediction using neural networks and deep learning: A review. *Comput. Biol. Chem.* **2019**, *81*, 1–8.