Computational Approaches for Structure-Based Molecular Characterization and Functional Annotation of the Fusion Protein of *Nipah henipavirus* †

Abu Saim Mohammad Saikat 1,* , Ranjit Chandra Das 2 and Madhab Chandra Das 3

1 Department of Biochemistry and Molecular Biology, Life Science Faculty, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj 8100, Bangladesh
2 Department of Materials Science and Engineering, Florida State University, Tallahassee, FL 32306, USA
3 Department of Electrical & Electronic Engineering, University of Information Technology and Sciences (UITS), Dhaka 1212, Bangladesh
*
Correspondence: asmsaikat.bmb@gmail.com; Tel.: +880-1795840353

Abstract: Throughout history, viral epidemics of varying frequency and intensity have been responsible for inducing panic and causing widespread damage. The Nipah virus has one of the highest rates of fatalities of any infectious disease in the world. There have been cases when severe respiratory distress has resulted in death, and it is known that these cases can cause encephalitis. The appearance of the virus and its ability to spread are affected by several factors. Several strategies have been created to raise awareness about the need for personal hygiene and enhance surveillance within the contaminated zone. This work aimed to determine the characteristics of a previously unidentified protein linked with the fusion of *Nipah henipavirus* particles. The protein’s secondary structure comprises helix, sheet, turn, and secondary coil structures. The protein is a fusion protein. In addition, the estimated Ramachandran plot provided evidence of the accuracy of the modeled protein structure. This accuracy was then verified by the Z-score-based and local model quality evaluation methods. It is possible to think of the protein as a target for developing prospective therapeutic and vaccine candidates directed against the protein to fight viral infections.

Keywords: Nipah virus; protein characterization; Protein stability; GRAVY; viral infection

1. Introduction

The Nipah virus (NiV), spread by bats and capable of causing fatal encephalitis in humans, 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 received 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 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 the Nipah (NiV) and Hendra (HeV) viruses. This results in encephalitis, pneumonia, and frequent deaths. 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*, which is 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 for 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 the 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. Additionally, 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*.

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 the left and right hands [31–33]. The ExPASy ProtParam tool identified the physicochemical characteristics of the protein, such as amino acid compositions, atomic compositions, and protein half-life calculations (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%), followed by oxygen (817, 9.5%), nitrogen (693, 8.1%), and sulfur (26, 0.3%).

Table 1. Protein retrieval.

<table>
<thead>
<tr>
<th>Protein Individualities</th>
<th>Protein Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Locus</td>
<td>QBQ56722</td>
</tr>
<tr>
<td>Amino acid</td>
<td>546 aa</td>
</tr>
<tr>
<td>Accession</td>
<td>QBQ56722</td>
</tr>
<tr>
<td>Version</td>
<td>QBQ56722.1</td>
</tr>
<tr>
<td>GenBank ID</td>
<td>QBQ56722.1</td>
</tr>
<tr>
<td>Source</td>
<td>Nipah henipavirus</td>
</tr>
<tr>
<td>Organism</td>
<td>Nipah henipavirus</td>
</tr>
</tbody>
</table>

FASTA sequence

>QBQ56722.1 fusion protein [Nipah henipavirus]
MAVLNKRKYSNLLLILMISECSVGLHYEKLSKIGLVKG
ITRKYKKSNPITKDIVKIMIPNVSNSMQCTQGSVMENYK
TRLGILITPIKGALEIYKNNTHDIVGDYRLAGVIMAEG
AIGIATAAQITAGVALYAMEKNADNKKSEISTENG
AVKLQGETAETVYVLTAQDÍYINTLVTIDKISKQTE
LSDLALSKYLSDLLFVFGPNLQDPVSNSMTIQAIASQAFA
GNYETLRTLGLAYTEDDDLESDDTGFQIQV/DLSGYNYIV
RYVFPILTEQQAYIQELLIPSIFNNDNSEISIVPNFILYRN
TLISNIEIGPCLTIRSKVCNQDYATPMTNMRECLKSTG
EKCPRELVSSHVPFRALSNGVLFANCISVTQCQQTIGRA
ISQSGEQTLMLNMDTTCPTAVLCNIVSLKLYGSLNYNS
EGIAIGPPVFTDNVDDISOSSMNQSLQSDKYKEAQ
RLLDTVNPLSILMSLMIYVLSIASLCIGLITFSFIIVEKKRN
TYSRLEDRRVPRPTSSGDLVYIGT

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 (1, 0.2%), 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 60,280.90 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 the development of numerous biotherapeutics is now standard practice. Various options are available for fine-tuning the half-life and adaptation to the desired treatment method and condition [39–42]. The anticipated protein half-life is 30 h (mammalian reticulocytes, in vitro); >20 h (yeast, in vivo); and >10 h (Escherichia coli, in vivo).

Table 2. Physicochemical parameters of the selected protein.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecular weight</td>
<td>60,280.90 Da</td>
</tr>
<tr>
<td>Theoretical pl</td>
<td>6.08 (6.30 *)</td>
</tr>
<tr>
<td>Total number of positively charged</td>
<td></td>
</tr>
<tr>
<td>residues (Arg + Lys)</td>
<td>46</td>
</tr>
<tr>
<td>Total number of negatively charged</td>
<td></td>
</tr>
<tr>
<td>residues (Asp + Glu)</td>
<td>48</td>
</tr>
<tr>
<td>Total number of atoms</td>
<td>8584</td>
</tr>
<tr>
<td>Estimated half-life</td>
<td></td>
</tr>
<tr>
<td>(a) 30 h (mammalian reticulocytes, in vitro)</td>
<td></td>
</tr>
<tr>
<td>(b) &gt;20 h (yeast, in vivo)</td>
<td></td>
</tr>
<tr>
<td>(c) &gt;10 h (Escherichia coli, in vivo)</td>
<td></td>
</tr>
<tr>
<td>Instability index (II)</td>
<td>38.05</td>
</tr>
<tr>
<td>Aliphatic index</td>
<td>112.27</td>
</tr>
<tr>
<td>Grand average of hydropathicity (GRAVY)</td>
<td>0.177</td>
</tr>
</tbody>
</table>

* pI calculated by the SMS v2.0 tool.

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 with an 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.

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 in 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 an alpha helix (239, 43.77%), an extended strand (112, 0.51%), a beta turn (23, 4.21%), and a 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.
Instability index (II) 38.05
Aliphatic index 112.27
Grand average of hydropathicity (GRAVY) 0.177
*pI calculated by the SMS v2.0 tool.

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 with an 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 [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.

### 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 in 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 an alpha helix (239, 43.77%), an extended strand (112, 0.51%), a beta turn (23, 4.21%), and a 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 (purple color).

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

![Amino acid types of the selected protein](image)

![Sequence plot of the selected protein](image)
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 template (HHpred ID: 2B9B_A) was selected for anticipation of the three-dimensional protein structure by the Modeller program with the HHpred interface with a probability of 100%, an E-value of $2.8 \times 10^{-132}$, and a target length of 497 (Figure 5).

![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.](image)

The estimated Ramachandran plot calculations of the selected protein were as follows: 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.
4. Conclusions

NiV has developed into 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 comprehensive 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.

Author Contributions: Conceptualization, A.S.M.S.; methodology, A.S.M.S.; software, A.S.M.S.; validation, A.S.M.S.; formal analysis, A.S.M.S.; investigation, A.S.M.S.; resources, A.S.M.S.; formal analysis, A.S.M.S.; investigation, A.S.M.S.; resources, A.S.M.S.; writing—original draft preparation, A.S.M.S.; writing—review and editing, A.S.M.S., R.C.D. and M.C.D.; visualization, A.S.M.S.; 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.

Acknowledgments: The author would like to acknowledge the preprint version of the manuscript available at ChemRxiv (DOI:10.26434/chemrxiv-2022-mcm52).

Conflicts of Interest: The author declares no conflict of interest.

References


2. Rahman, M.; Chakraborty, A. Chakraborty, Nipah virus outbreaks in Bangladesh: A deadly infectious disease. WHO South East Asia J. Public Health 2012, 1, 208–212. [CrossRef]


