



ANTIGENIC DIVERSITY IN THE IMMUNODOMINANT PROTEIN OF NEWCASTLE DISEASE VIRUS

Gurupriya. K

M.Sc. Bioinformatics, Bioscience Department, Sri Krishna Arts and Science College, Coimbatore, Tamil Nadu, India ¹

Prof J. Vasanth Nirmal Bosco, Department of Bioscience, Sri Krishna Arts and Science College, Coimbatore, Tamil Nadu, India ¹

Dr B. Aarthi Rashmi, Department of Bioscience, Sri Krishna Arts and Science College, Coimbatore, Tamil Nadu, India ²

Dr R. Venkataramanam, PhD, Professor and Head. Bioinformatics Centre, Madras Veterinary College, Chennai ³

Corresponding author email ID: hodbitmvc@tanuvas.org.in, gurupriyakasi@gmail.com.

Abstract: *Newcastle Disease Virus (NDV) is a highly contagious viral pathogen affecting domestic and wild birds, with major importance in poultry production. The Hemagglutinin-Neuraminidase (HN) protein is an important surface and immunodominant protein of NDV because it participates in receptor binding, neuraminidase activity, viral attachment and release. The present internship study was undertaken to investigate the antigenic diversity of the HN protein using a combination of sequence analysis, protein structure prediction, structural visualisation, structure validation and phylogenetic analysis. HN protein sequences from selected NDV isolates were retrieved from the NCBI GenBank database in FASTA format. Homologous sequences and conserved regions were examined using BLAST. The three-dimensional structure of the HN protein was obtained from the AlphaFold Protein Structure Database and visualised using PyMOL. Structural features including alpha-helices, beta-sheets, loops, surface-exposed regions and flexible segments were examined. BIOVIA Discovery Studio*

was used for structural validation through Ramachandran plot analysis and for assessing the stereochemical quality of the predicted model. Multiple sequence alignment was performed using ClustalW in MEGA, followed by phylogenetic analysis to investigate sequence similarity, genetic divergence and clustering among selected isolates. The structural model showed an organised arrangement of secondary structural elements with exposed and flexible regions. The Ramachandran analysis indicated that most residues were located in favoured and additionally allowed regions, supporting acceptable structural quality. Phylogenetic analysis revealed distinct clusters and genetic diversity among the selected isolates, with closely related Indian isolates forming related groups. The combined analysis demonstrates how sequence, structural and evolutionary approaches can be used to characterise HN protein diversity and to support molecular surveillance and future vaccine-related investigations of Newcastle Disease Virus.



Keywords: Newcastle Disease Virus, HN protein, antigenic diversity, immunodominant protein, GenBank, BLAST, AlphaFold, PyMOL, BIOVIA Discovery Studio, Ramachandran plot, MEGA, ClustalW, phylogenetic analysis

1. INTRODUCTION

Newcastle disease is a highly contagious viral disease that affects a wide range of bird species, particularly domestic poultry such as chickens, turkeys, ducks and pigeons. The disease is caused by Newcastle Disease Virus (NDV) and may produce respiratory problems, neurological manifestations such as tremors and paralysis, reduction in egg production, greenish diarrhoea and, in severe cases, high mortality. The rapid spread of the infection creates an important challenge for poultry health and production.

Transmission can occur through direct contact with infected birds and indirectly through contaminated feed, water, aerosols, equipment, surfaces and clothing. Because the disease can spread rapidly and can be associated with economic losses and restrictions on poultry movement and trade, understanding the molecular characteristics of NDV is important for disease surveillance and control

1.1 Newcastle Disease Virus

Newcastle Disease Virus is an enveloped virus with a negative-sense single-stranded RNA genome and belongs to the Paramyxoviridae family. NDV strains can differ in their level of virulence and are commonly described as lentogenic, mesogenic and velogenic types. The diversity in virulence and genetic composition contributes to differences in disease presentation and creates difficulties for consistent disease control.

The viral envelope contains important proteins involved in infection. Among them, the Hemagglutinin-Neuraminidase (HN) and Fusion (F) proteins have major roles in

attachment and entry. HN participates in receptor binding and neuraminidase activity, while F is involved in fusion of the viral envelope with the host-cell membrane. These functions make the surface proteins important targets for molecular and immunological studies.

1.2 Genome and Important Viral Proteins

The NDV genome encodes structural and non-structural proteins that contribute to viral replication, pathogenicity, host interaction and immune evasion. Molecular studies have particularly emphasised proteins such as HN, F, P, M and V. The HN and F proteins are especially relevant to viral attachment and membrane fusion. The V protein has also been associated with host-range restriction and interference with innate immune responses.

Variation in genes encoding important viral proteins can influence biological properties. In particular, sequence changes in HN and F are relevant when studying genetic diversity, virulence and possible changes in antigenic properties. Therefore, sequence-based investigation of these proteins is useful for understanding NDV evolution.

1.3 Hemagglutinin-Neuraminidase (HN) Protein

The Hemagglutinin-Neuraminidase protein is a multifunctional glycoprotein located on the surface of NDV. It possesses both hemagglutination and neuraminidase activities. These activities contribute to interaction with host-cell receptors and to the release of newly formed viral particles. The protein contains antigenic regions and epitopes that can be recognised by the host immune system.

Because the HN protein is surface-exposed, changes in its amino-acid sequence can be relevant to antigenic characteristics. Structural regions such as loops and surface-exposed segments can show flexibility and may be important when interpreting sequence variation. For these reasons, HN was selected as the major protein of interest in this internship study.



1.4 Immunodominant Proteins

Immunodominant proteins are viral proteins that strongly stimulate host immune responses and contribute to the generation of antibodies. In NDV, the HN and F proteins are important immunodominant proteins. HN is particularly relevant because it is exposed on the viral surface and directly participates in host-cell receptor interaction.

Variation in an immunodominant protein may influence recognition by the host immune system. Therefore, examining conserved and variable regions of HN provides a computational approach for investigating potential antigenic diversity among different NDV isolates.

1.5 Antigenic Diversity

Antigenic diversity refers to differences in viral antigens that can be recognised by the host immune system. In RNA viruses, sequence variation can arise during replication, and mutations in viral proteins can lead to amino-acid substitutions. When such substitutions occur in surface proteins, they may alter structural features or antigenic characteristics.

For NDV, antigenic diversity is relevant to repeated outbreaks and to the evaluation of vaccine effectiveness. Continuous molecular characterisation can therefore help identify conserved regions as well as variable regions among circulating isolates. The present study uses HN sequence and structure analysis to examine these differences.

1.6 Role of Sequence Variation in Antigenic Diversity

Sequence comparison provides a direct way of identifying similarities and differences among protein isolates. Conserved residues or regions can indicate functional or structural constraints, whereas variable positions may represent regions that have undergone greater sequence change. Multiple sequence alignment is particularly useful for displaying these patterns across a set of sequences.

In this study, HN protein sequences were compared using sequence-analysis workflows and subsequently examined in a phylogenetic framework. The resulting patterns were interpreted together with the predicted protein structure rather than using sequence differences alone.

1.8 Protein Modelling and Structural Analysis

Protein modelling is a computational method for obtaining a three-dimensional representation of a protein from its amino-acid sequence. The AlphaFold Protein Structure Database provides predicted structures that can be used for downstream visualisation and analysis. In the present work, the HN model was obtained and downloaded in PDB format.

PyMOL was used to visualise the predicted structure. Different representations allowed examination of the overall fold, secondary structural elements, loops, surface-exposed regions and flexible segments. Structural analysis helps provide a spatial context for interpreting sequence variability.

1.8 Protein Structure Validation

Predicted protein structures require validation to assess whether the model has acceptable stereochemical properties. The Ramachandran plot is a commonly used validation method that displays the distribution of amino-acid residues according to their backbone phi and psi torsion angles. Residues in favoured and allowed regions generally support acceptable backbone geometry, whereas residues in disallowed regions require closer examination.

In this internship, BIOVIA Discovery Studio was used for Ramachandran plot analysis. The validation results were considered together with PyMOL observations of missing residues and flexible regions.

1.9 Phylogenetic Analysis

Phylogenetic analysis investigates evolutionary relationships among sequences. In NDV research, aligned



HN or F sequences can be used to examine clustering, genetic divergence and relationships among isolates. Multiple sequence alignment creates a common positional framework, after which a phylogenetic tree can be constructed using MEGA.

The tree generated in this study was interpreted in terms of clustering patterns and sequence similarity. Bootstrap analysis was used to assess the reliability of tree topology. Closely related isolates were observed as related clusters, while differences between clusters indicated genetic divergence.

1.10 Importance of Computational Analysis in NDV Research

The combination of sequence retrieval, homology analysis, structural prediction, visualization, structural validation and phylogenetic analysis provides a multi-level computational framework. Sequence methods identify variation, structural methods provide three-dimensional context, and phylogenetic analysis places the observed sequence relationships in an evolutionary framework. Such integrated analysis is useful for documenting molecular variation in important viral proteins.

2. REVIEW OF LITERATURE

2.1 Avian Immune Response to NDV

Seal, King and Sellers described the avian response to Newcastle disease virus and emphasised the importance of host immune mechanisms in controlling infection. Kapczynski, Afonso and Miller further discussed innate and adaptive immune responses, including interferon-related responses, antibodies and T-cell responses. These studies provide the immunological background for examining viral proteins that interact with host immunity.

2.2 HN Protein and Viral Tropism

Huang and colleagues investigated the HN protein of NDV and demonstrated its importance in determining viral

tropism and virulence. The multifunctional role of HN in receptor binding, neuraminidase activity and promotion of fusion highlights why sequence and structural variation in HN is biologically relevant.

2.3 HN Protein Length and Biological Activity

Jin and colleagues examined HN protein length diversity and reported that differences in HN length can influence viral replication, virulence and biological activities. This work supports the importance of examining sequence variation when comparing NDV isolates.

2.4 HN Gene Sequence Variation

Millar, Chambers and Emmerson analysed the nucleotide sequence of the HN gene and showed that sequence information can provide insight into viral evolution and variation. Ke and colleagues also examined HN sequence and haemagglutinin activity, identifying sequence differences associated with functional characteristics.

2.5 HN Protein and Apoptosis

Ravindra and colleagues investigated the HN protein in chicken embryo fibroblast cells and reported an association between HN expression and apoptosis. These observations demonstrate that HN can participate in host-cell effects beyond receptor interaction.

2.6 NDV V Protein and Host Range

Park and colleagues studied the V protein of NDV and its role in host-range restriction. The study highlighted the importance of viral proteins that interfere with host innate immune responses and contribute to host adaptation.

2.7 Recombinant NDV and Vaccine Applications

Veits and colleagues developed recombinant NDV expressing the H5 hemagglutinin gene and reported protection against both Newcastle disease and avian influenza. Such work illustrates the potential of NDV as a viral-vector platform and emphasizes the importance of understanding viral surface proteins in vaccine research.



2.8 C-Terminal Extension of HN

Zhao and colleagues investigated the C-terminal extension of the HN gene. Their studies examined whether variation in this region influences virulence, biological activity and enteric tropism. These investigations contribute to understanding how structural regions of HN relate to viral properties.

2.9 Genomic Variation and Vaccine Effectiveness

Kattenbelt and colleagues examined sequence variation in the NDV genome and emphasized that changes in important viral genes can contribute to differences in virulence and antigenic characteristics. This provides a basis for molecular surveillance and evaluation of vaccine-related challenges.

2.10 Structural Studies of the HN Ectodomain

Yuan and colleagues reported structural information for the NDV HN ectodomain and described a four-helix bundle stalk involved in interaction with the fusion protein. Structural studies of this type provide a framework for interpreting predicted models and for relating protein regions to viral functions.

2.11 Evolutionary Dynamics of NDV

Miller and colleagues described the evolutionary dynamics of Newcastle disease virus. Their work supports the use of phylogenetic analysis to understand genetic diversity, lineage relationships and changes occurring in circulating NDV populations.

2.12 Recent Perspectives on NDV and Poultry Productivity

Kamal and colleagues reported relationships between Newcastle disease and laying-hen productivity, including effects on egg production and flock performance. Bello and colleagues reviewed current and emerging diagnostic and vaccination approaches. Together, these studies emphasize

the practical importance of continued molecular and epidemiological investigation of NDV.

3. OBJECTIVES

3.1 Primary Objective

To analyse the antigenic diversity of the Hemagglutinin-Neuraminidase (HN) protein of Newcastle Disease Virus using computational bioinformatics approaches.

3.2 Specific Objectives

- To retrieve HN protein sequences from selected Newcastle Disease Virus isolates.
- To compare HN sequences and identify conserved and variable regions.
- To obtain and visualize the predicted three-dimensional structure of the HN protein.
- To examine missing residues, flexible regions and structural organization using PyMOL.
- To validate the predicted protein model using Ramachandran plot analysis in BIOVIA Discovery Studio.
- To perform multiple sequence alignment using ClustalW.
- To construct and interpret a phylogenetic tree using MEGA.
- To relate sequence variation, structural features and phylogenetic clustering to the study of HN protein diversity.

4. MATERIALS AND METHODS

4.1 Overall Computational Workflow

The study followed an integrated computational workflow beginning with retrieval of HN protein sequences from NCBI GenBank. The selected sequences were examined using BLAST, and the sequences were prepared for comparative analysis. Protein structure information was obtained from AlphaFold and visualized in PyMOL. Structural validation was performed using BIOVIA



Discovery Studio. Multiple sequence alignment was performed using ClustalW in MEGA, followed by phylogenetic tree construction and interpretation.

4.2 Sequence Data Retrieval from NCBI GenBank

NCBI GenBank was used as the primary sequence resource. The database contains annotated nucleotide and protein sequences submitted by researchers. HN protein sequences of selected NDV isolates were retrieved in FASTA format. The downloaded sequences provided the input for subsequent sequence comparison, alignment and phylogenetic analysis.

4.3 BLAST-Based Homology Analysis

NCBI BLAST was used to identify sequences homologous to the selected HN protein sequence and to examine sequence similarity. Homology information helps confirm that the retrieved sequences correspond to related HN proteins and supports the selection of sequences for comparative analysis.

4.4 AlphaFold-Based Protein Structure Prediction

The AlphaFold Protein Structure Database was used to obtain a predicted three-dimensional structure for the HN protein. The relevant protein sequence or accession information was used to locate the available predicted model. The structure was downloaded for further visualization and validation.

4.5 PDB Structure Retrieval

The predicted HN structure was downloaded in PDB format. This structure file provided the coordinate information required for visualization and downstream structural analysis using molecular graphics and validation software.

4.6 PyMOL-Based Structural Visualization

PyMOL was used to visualize the HN protein in three dimensions. Cartoon and surface representations were used to inspect the protein fold, alpha-helices, beta-sheets,

loops, exposed regions and flexible segments. The software also helped in examining discontinuities and missing residues within the predicted structure.

4.7 BIOVIA Discovery Studio Analysis

BIOVIA Discovery Studio was used to examine the structural model and to perform validation. The software provided a graphical environment for structural inspection and analysis of the predicted HN protein.

4.8 Ramachandran Plot Validation

Ramachandran plot analysis was performed to evaluate the stereochemical quality of the predicted model. Residues were assessed according to their backbone conformational angles. The distribution in favoured, additionally allowed, and disallowed regions was considered when interpreting the reliability of the structure.

4.9 Multiple Sequence Alignment Using ClustalW

Selected HN protein sequences were imported into MEGA and aligned using ClustalW. Alignment enabled identification of conserved positions and variable regions across the sequences. The aligned data were then used as the basis for phylogenetic analysis.

4.10 Phylogenetic Tree Construction Using MEGA

MEGA was used to construct a phylogenetic tree from the aligned HN sequences. The tree was examined for clustering patterns, sequence similarity and genetic divergence among the selected NDV isolates.

4.11 Bootstrap-Based Tree Reliability

Bootstrap analysis was incorporated to assess the reliability of the observed phylogenetic clustering. Higher support for a branch indicates stronger consistency of that grouping within the analysis. The tree was interpreted together with the sequence relationships rather than as an independent measure of antigenicity.

5. RESULTS AND DISCUSSION



5.1 Protein Structure Prediction

The three-dimensional structure of the NDV HN protein was successfully obtained from the AlphaFold Protein Structure Database. The model displayed an organised conformational arrangement containing alpha-helices, beta-sheets, loops and coil regions. The predicted model provided a structural framework for examining the HN protein beyond its linear sequence.

5.2 Three-Dimensional Structural Organisation of HN

Visualisation of the predicted HN model showed a compact overall fold. The ribbon representation allowed identification of secondary structural elements and the spatial organisation of the protein. Surface-exposed regions were particularly relevant to interpretation because the HN protein functions at the viral surface.

5.3 Surface-Exposed and Flexible Regions

Several exposed loops and flexible regions were observed in the predicted structure. Such regions are structurally different from the more stable core and can be important when interpreting sequence variation. The analysis also identified regions with comparatively lower confidence or greater flexibility in the model.

5.4 Missing Residue Analysis

PyMOL analysis revealed discontinuities and missing residues mainly in flexible loops and terminal regions. The missing or discontinuous regions were not predominantly located in the stable alpha-helical and beta-sheet core. This observation was considered when interpreting the predicted model and its structural reliability.

5.5 Ramachandran Plot Analysis

Ramachandran plot analysis using BIOVIA Discovery Studio showed that the majority of amino-acid residues occurred in favoured and additionally allowed regions. Only a small number of residues were located in disallowed regions, with these mainly associated with flexible loops or terminal

segments. The distribution supported acceptable stereochemical quality and stability of the predicted HN model.

5.6 Multiple Sequence Alignment

Multiple sequence alignment using ClustalW provided a comparative view of HN protein sequences from the selected NDV isolates. Conserved regions were retained across related sequences, whereas variable positions reflected sequence differences among isolates. The aligned sequences formed the basis for the subsequent evolutionary analysis.

5.7 Phylogenetic Clustering

The phylogenetic analysis produced distinct clusters among the selected HN sequences. Closely related sequences were grouped together, while sequences showing greater divergence occurred in separate branches. Indian isolates included in the analysis showed closely related clustering patterns, suggesting regional evolutionary relationships within the selected dataset.

5.8 Genetic Diversity Among NDV Isolates

The branching pattern demonstrated genetic diversity among the analysed HN sequences. Differences between clusters indicated sequence divergence and evolutionary separation. Bootstrap support was used to assess the consistency of the observed tree topology.

5.9 Antigenic Diversity and Structural Variation

The study combines sequence and structure observations to understand HN diversity. Variable sequence regions can be considered alongside exposed and flexible structural regions when interpreting potential antigenic variation. However, the computational analysis performed in this internship identifies sequence and structural differences and does not by itself experimentally establish antibody escape or altered vaccine protection.

5.10 Overall Interpretation



The combined results indicate that the HN protein contains a balance of structurally conserved regions and variable surface-associated segments. Structural validation supports the usability of the predicted model, while phylogenetic clustering demonstrates genetic relationships among the selected isolates. Together, these findings provide a computational perspective on NDV HN protein diversity.

5.11 Figure and Result Presentation

Figure 1. Structure of HN protein obtained from the predicted structural model.

Figure 2. Missing residue analysis of HN protein using PyMOL.

Figure 3. Ramachandran plot of HN protein generated using BIOVIA Discovery Studio.

Figure 4. Phylogenetic tree of HN protein sequences constructed using MEGA.

3. CONCLUSIONS

The integrated computational analysis demonstrates that the HN protein of Newcastle Disease Virus contains both conserved structural characteristics and sequence variation among isolates. The AlphaFold model provided a useful three-dimensional representation, while PyMOL and BIOVIA Discovery Studio enabled structural examination and validation. Most residues occupied favoured and allowed Ramachandran regions, indicating acceptable model quality. Phylogenetic analysis further demonstrated genetic relationships and diversity among the selected isolates. Overall, the study supports the use of combined bioinformatics approaches to characterise NDV HN protein variation and provides a foundation for continued molecular surveillance and future studies on antigenicity and vaccine development.

ACKNOWLEDGEMENT

The authors can acknowledge any person/authorities in this section. This is not mandatory.

REFERENCES

- [1] Bello, M.B., Yusoff, K., Ideris, A., Hair-Bejo, M., Peeters, B.P.H., & Omar, A.R. (2018). Diagnostic and Vaccination Approaches for Newcastle Disease Virus in Poultry: The Current and Emerging Perspectives. *BioMed Research International*, 2018, 1–18. DOI: 10.1155/2018/7278459.
- [2] Huang, Z., Panda, A., Elankumaran, S., Govindarajan, D., Rockemann, D.D., & Samal, S.K. (2004). The Hemagglutinin-Neuraminidase Protein of Newcastle Disease Virus Determines Tropism and Virulence. *Journal of Virology*, 78(8), 4176–4184.
- [3] Jin, J., Zhao, J., Ren, Y., Zhong, Q., & Zhang, G. (2016). Contribution of HN protein length diversity to Newcastle disease virus virulence, replication and biological activities. *Scientific Reports*, 6, 36890.
- [4] Kamal, M.A.M., Atef, M., Khalf, M.A., & Ahmed, Z.A.M. (2024). Newcastle viral disease causation web correlations with laying hen productivity. *Scientific Reports*, 14, 16021.
- [5] Kapczynski, D.R., Afonso, C.L., & Miller, P.J. (2013). Immune responses of poultry to Newcastle disease virus. *Developmental & Comparative Immunology*, 41(3), 447–453.
- [6] Kattenbelt, J.A., Stevens, M.P., & Gould, A.R. (2006). Sequence variation in the Newcastle disease virus genome. *Virus Research*, 116(1–2), 168–184.
- [7] Ke, G.M., Chuang, K.P., Chang, C.D., Lin, M.Y., & Liu, H.J. (2010). Analysis of sequence and haemagglutinin activity of the HN glycoprotein of Newcastle disease virus. *Avian Pathology*, 39(3), 235–244.
- [8] Millar, N.S., Chambers, P., & Emmerson, P.T. (1986). Nucleotide Sequence Analysis of the Haemagglutinin-Neuraminidase Gene of Newcastle Disease Virus. *Journal of General Virology*, 67(9), 1917–1927.
- [9] Miller, P.J., Kim, L.M., Ip, H.S., & Afonso, C.L. (2009). Evolutionary dynamics of Newcastle disease virus. *Virology*, 391(1), 64–72.



- [10] Nagai, Y., Hamaguchi, M., & Toyoda, T. (1989). Molecular biology of Newcastle disease virus. *Progress in Veterinary Microbiology and Immunology*, 5, 16–64.
- [11] Park, M.S., García-Sastre, A., Cros, J.F., Basler, C.F., & Palese, P. (2003). Newcastle Disease Virus V Protein Is a Determinant of Host Range Restriction. *Journal of Virology*, 77(17), 9522–9532.
- [12] Ravindra, P.V., et al. (2008). HN protein of Newcastle disease virus causes apoptosis in chicken embryo fibroblast cells. *Archives of Virology*, 153(4), 749–754.
- [13] Seal, B.S., King, D.J., & Sellers, H.S. (2000). The avian response to Newcastle disease virus. *Developmental & Comparative Immunology*, 24(2–3), 257–268.
- [14] Shah, M., Bharadwaj, M.S.K., Gupta, A., Kumar, R., & Kumar, S. (2019). Chicken viperin inhibits Newcastle disease virus infection in vitro: A possible interaction with the viral matrix protein. *Cytokine*, 120, 28–40.
- [15] Sun, W., et al. (2020). Newcastle disease virus expressing the spike protein of SARS-CoV-2 as a live virus vaccine candidate. *EBioMedicine*, 62, 103132.
- [16] Veits, J., et al. (2006). Newcastle disease virus expressing H5 hemagglutinin gene protects chickens against Newcastle disease and avian influenza. *Proceedings of the National Academy of Sciences*, 103(21), 8197–8202.
- [17] Yuan, P., Leser, G.P., Demeler, B., Lamb, R.A., & Jardetzky, T.S. (2008). Structure of the Newcastle disease virus hemagglutinin-neuraminidase (HN) ectodomain reveals a four-helix bundle stalk. *Proceedings of the National Academy of Sciences*, 105(39), 14920–14925.
- [18] Zhao, W., Hu, H., Zsak, L., Yu, Q., & Yang, Z. (2013). HN gene C-terminal extension of Newcastle disease virus is not the determinant of the enteric tropism. *Virus Genes*, 47(1), 27–33.
- [19] Zhao, W., Zhang, Z., Zsak, L., & Yu, Q. (2013). Effects of the HN gene C-terminal extensions on the Newcastle disease virus virulence. *Virus Genes*, 47(3), 498–504.