ISSN 2348-1218 (print)
International Journal of Interdisciplinary Research and Innovations ISSN 2348-1226 (online)
Vol. 5, Issue 1, pp: (19-24), Month: January - March 2017, Available at: www.researchpublish.com

Protein Sequence Analysis of Newcastle (*Paramyxovirus*) in Poultry

[1]Ayuba. Dauda, [2]Haruna Duwa, [1]Jummai Adamu Yaska, [1]Joseph Ochoche Egahi, [1]Yarima Shettima Kollo

[1] University of Agriculture Makurdi P.M.B 2373, Makurdi, Nigeria
[2]University of Maiduguri P.M.B 1069,Maiduguri, Nigeria.

Abstract: **A total of ten (10) Newcastle proteins of poultry were retrieved from the GenBank (www.ncbi.nlm. nih.gov). The Genbank accession numbers of the sequences and sequence variations were used to investigate the molecular identity of various newcastle proteins. The phyco-chemical properties of newcastle proteins were performed using protparam tool. The isoelectric point (pI), extinction coefficient (EC); instability index (II), aliphatic index (AI) and grand average of hydropathicity (GRAVY) were also computed. The study reveals that 90% and 10% of the pI of Newcastle protein were basic nature and acidic in nature respectively. The EC and II of Newcastle protein shows better stability which is an indication of resistance to mutation. AI for all the protein is <100. This indicates that the newcastle protein are not thermally stable. The GRAVY, 90% of the protein were negative and only 10% were positive. The positive values indicate solubility in water while negative is not soluble in water. The amino acid composition of newcastle proteins indicate that they are rich in serine and asparagines which are hydroxyl amino acid which is non reactive and can play a role in substrate recognition which mean they are resistant to mutation. The prediction of secondary structure was performed using SOPMA. The proteins are more of random coil structure then followed by alpha helix which shows that they are not thermally stable. Phyre2 server was used to predict the 3D structure of newcastle proteins. Study on pharmacogenetic and mutagenesis is needed regarding this finding.**

Keywords: **Protein, Newcastle and Sequence.**

1. INTRODUCTION

Newcastle disease (ND) is a deadly viral disease of poultry all over the world since the time of its first isolation in England in 1926 till today (Banu *et al.*, 2009). It is considered as one of the major economic threats to poultry population because of the high morbidity, which may vary from 90 to 100% in unprotected birds depending on the virulence of the Newcastle disease virus (NDV) exposed (Alexander, 2003). The causative agent, NDV is also known as Avian *Paramyxovirus* type 1(APMV-1) and the only member of the genus of *Avulavirus* (Mayo, 2002) in the family *Paramyxoviridae*. In most developing countries, ND ranked as the most important poultry disease affecting the poultry population with consequent great economic losses (Aini, 1990; Martin, 1992). In Nigeria with the outbreak of this deadly disease had cause a serious loss in poultry industry. Although much veterinary work had been done on this disease but little or no effort has been made genetically in Nigeria. The aim of this study is to carry out protein sequence analysis in order to provide genetic information which may help to curtail the effect of this deadly disease of poultry

2. MATERIAL AND METHODS

A total of ten (10) newcastly proteins of poultry were retrieved from the GenBank (www.ncbi.nlm.nih.gov). The Genbank accession numbers of the sequences and sequence variations are shown in Tables 1

ISSN 2348-1218 (print)
International Journal of Interdisciplinary Research and Innovations ISSN 2348-1226 (online)
Vol. 5, Issue 1, pp: (19-24), Month: January - March 2017, Available at: www.researchpublish.com

Table 1: Protein names, accession number and amino acid number

Protein Name	Accession Number	Amino acid Number
H10 Avian Haemmaglutinin	4CZ0	318
H5 Haemagglutinin	1JSN	325
H2 Avian Hemagglutinin	2WR2	509
Glycoprotein	CDN30041	389
fusion protein	CDN30037	538
Nucleoprotein	CDN30034	391
Glycoprotein [Avian metapneumovirus type C]	CDN30032	585
Fusion protein [Avian metapneumovirus type C	CDN30028	537
Haemagglutinin	4BH4	328
Influenza Virus Haemagglutinin	4BGY	326

ProtParam Tool was used for the computation of various physical and chemical properties of the newcastle proteins using amino acid sequences. The computed parameters were molecular weight, theoretical pI (isoelectric point), amino acid composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY) (Gasteiger, 2005). The amino acid sequences of Newcastle protein were subjected to secondary structure prediction using ExPASy's SOPMA tool. SOPMA is an improved SOPM method. It predicts 69.5% of amino acids for a 3 state description of the secondary structure (a helix, b sheets and coil). It predicts the secondary structure by consensus prediction from multiple alignments. The Phyre2 server was used to predict the 3D structure of CBPP and CCPP proteins. These servers predict the three-dimensional structure of a protein sequence using the principles and techniques of homology modeling (Kelley and Sternberg, 2009). Currently, the most powerful and accurate methods for detecting and aligning remotely related sequences rely on profiles or Hidden Markov Models (HMMs). 3DligandSite was used to predict the binding site of the 3D structure of the newcastle proteins. Phyre2 is coupled to the 3DligandSite server for protein binding site prediction (Wass *et al*., 2010).

3. RESULT AND DISCUSSIONS

The isoelectric point (pI), net charge (Q), extinction coefficient (EC), instability index (II), aliphatic index (AI) and grand average hydropathicity (GRAVY) of the protein are presented in table 2. All the protein are basic in nature (>7) only H2 Avian Hemagglutinin protein is acidic (5.84). The isoelectric point is of significance in protein purification because it is the pH at which solubility is always minimal and at which mobility in an electro focusing system is zero and therefore the point at which the protein will accumulate (Fennema, 2008). All the protein showed positive charge which mean they are extracellular protein. Only H2 Avian Hemagglutinin protein showed negative charge which mean is intracellular protein (Munduganore *et al*., 2012). The EC of a protein at 280 nm depends almost exclusively on the number of aromatic residues, particularly tryptophan (Gill *et al*., 1989). This indicates that the higher the EC value of the newcastle proteins, the higher the number of aromatic residues which made the protein highly stable (Gasteiger 2003; Munduganore *et al*., 2012). The II provides an estimate of the stability of protein in a test tube. A protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable (Guruprasad *et al.*, 1990). Glycoprotein fusion protein and Glycoprotein (Avian metapneumovirus type C) showed II > 40 which implies they are not stable while all the remaining protein selected for this study shows II <40 which implies they are stable and resistant to mutation from generation to generation. The AI of a protein is defined as the relative volume occupied by aliphatic side chains (alanine, valine, isoleucine, and leucine). It may be regarded as a positive factor for the increase of thermostability of globular proteins (Ikai, 1980). AI >100 indicates thermo stability (Munduganore *et al*., 2012). The AI in this study is <100 which implies they have less thermo stability. The **(**GRAVY) is an indication of protein solubility where a positive value correlated with hydrophobicity and negative value correlate with hydrophilicity. This study shows that most of the proteins are not soluble in water, only fusion protein (Avian metapneumovirus type C) is soluble in water. The amino acid

compositions of the Newcastle proteins are presented in table 3. The Newcastle proteins are high in serine, asparagini, glycine and leucine. Glycine and leucine are aliphatic residues Aliphatic side chains are very non-reactive, and are thus rarely involved directly in protein function, though they can play a role in substrate recognition. In particular, hydrophobic amino acids can be involved in binding/recognition of hydrophobic ligands such as lipids (Barnes *et al*., 1999). Asparagine is acidic which is quite frequently involved in protein-active or -binding sites. Serine and threonine are hydroxyls which are quite common in protein functional centres (Barnes *et al*., 1999). The hydroxyl group is fairly reactive, and can form hydrogen bonds with a variety of polar substrates (Barnes *et al*., 1999). Selenocystein and Pyrrolysine is zero for all the protein which a stop code (identity of the cannot be further determine). The predictions of secondary structure of Newcastle protein are presented in table 4. All the proteins show high value of random coil. Protein structure prediction from amino acid sequence is a fundamental scientific problem and it is regarded as a grand challenge in computational biology and chemistry. Given an amino acid sequence (i.e., the primary structure) which represents a monomeric globular protein in aqueous solution and at physiological temperatures, the objectives are to determine (i) all helical segments and all beta-strands, (ii) all pairs of beta-strands which form beta-sheets (i.e., the beta-sheet topology), (iii) all disulfide bridges if cysteines are present, (iv) all loops that connect secondary structure elements, and (v) the three-dimensional folded protein structure (Floudas, 2007). The practical applications of protein structure prediction are many and varied, including guiding the development of functional hypotheses about hypothetical proteins (Watson *et al*., 2005), improving phasing signals in crystallography (Qian *et al*., 2007), selecting sites for mutagenesis (Rava and Hussain, 2007) and the rational design of drugs (Park *et al*., 2008).

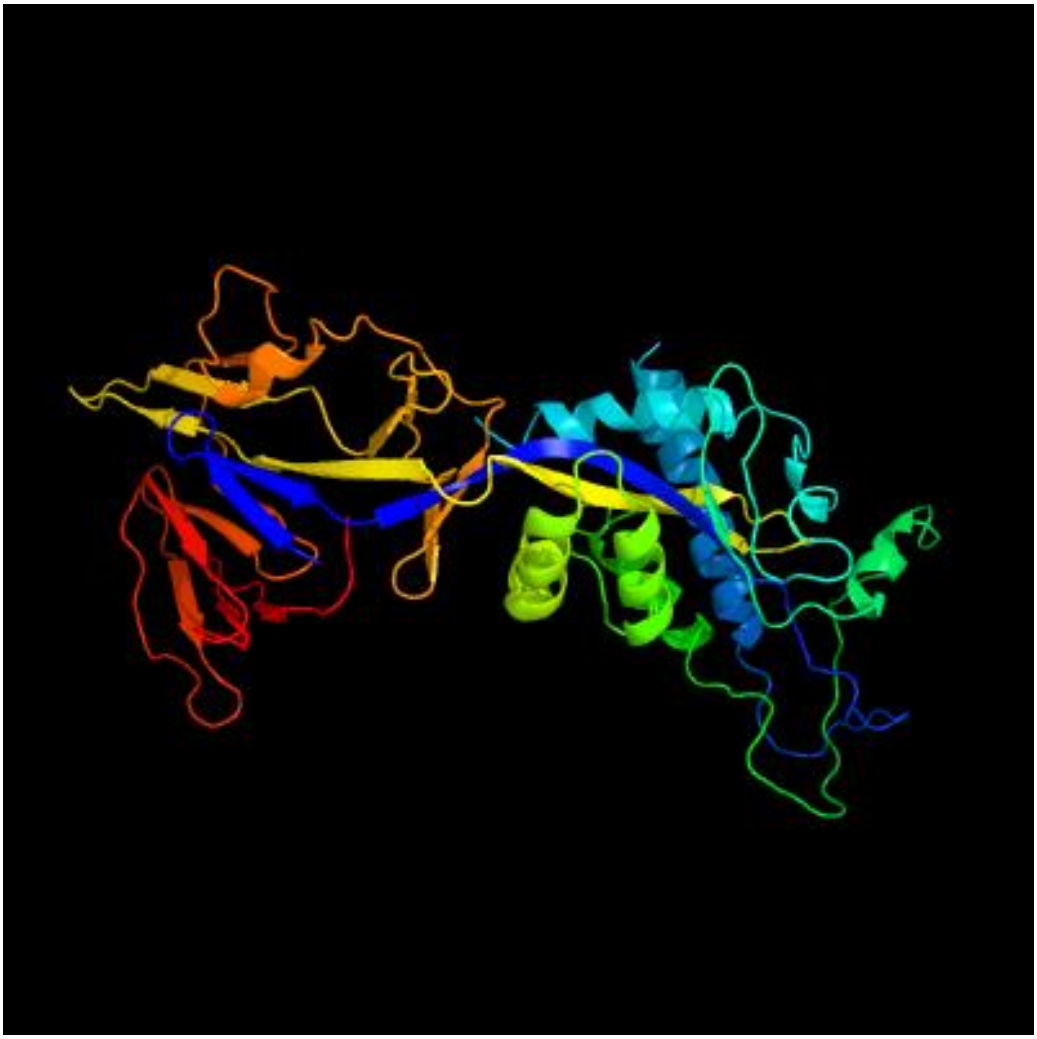

Image coloured by rainbow N → C terminus

Model dimensions (Å): **X**:88.960 **Y**:57.507 **Z**:67.585

Figure 1: Schematic 3D structure of H2 Avian Hemagglutinin of Newcastle proteins

ISSN 2348-1218 (print)
International Journal of Interdisciplinary Research and Innovations ISSN 2348-1226 (online)
Vol. 5, Issue 1, pp: (19-24), Month: January - March 2017, Available at: www.researchpublish.com

Table 2: Physico-chemical characteristic of Newcastle protein predicted by protparam

Protein	AA	Mol Wt	pI	Q	EC	Halflife	II	AI	GRAVY
H10 Avian Haemmaglutinin	318	34611.96	8.54	+ve	39420	1.1hr	37.25	72.36	-0.429
H5 Haemagglutinin	325	36531.2	7.70	+ve	52370	1.1hr	34.27	76.77	-0.490
H2 Avian Hemagglutinin	509	57525.0	5.84	_ve	83310	30hr	33.28	77.15	-0.488
Glycoprotein	389	41812.1	8.84	+ve	18910	30hr	46.80	58.46	-0.579
fusion protein	538	58494.9	8.24	+ve	41830	30hr	37.61	96.77	-0.015
Nucleoprotein	350	38253.2	8.86	+ve	25900	30hr	39.33	95.57	-0.002
Glycoprotein [Avian metapneumovirus type C]	585	63852.9	9.43	+ve	41370	30hr	42.15	33.13	-1.271
fusion protein [Avian metapneumovirus type C]	537	58075	7.82	+ve	42860	30hr	34.83	97.82	0.071
Haemagglutinin	328	37158.2	7.68	+ve	52370	1.1hr	31.29	80.82	-0.473
Influenza Virus Haemagglutinin	326	36906.8	7.69	+ve	52370	1.1hr	32.04	78.93	-0.487

AA=amino acid; pI=isoelectric point; Q=net charge; II=instability index; AI=alphatic index; GRAVY= grand average of hydropathicity; EC=extinctioncoefficient; Molwt=molecularweigh

Table 3: Amino Acid Composition (%) of Newcastle Protein

Protein	A	R	N	D	C	E	Q	G	H	I	L	K	M	F	P	S	T	W	Y	V	O	U
H10 Avian Haemmaglutinin	5.0	4.4	7.2	3.5	2.8	4.1	4.7	10.1	3.1	6.3	7.2	5.0	2.2	2.5	4.4	9.1	9.1	1.6	2.5	5.0	0.0	0.0
H5 Haemagglutinin	4.6	3.7	10.2	3.7	2.8	3.1	5.8	6.5	3.1	6.8	7.4	6.2	1.8	3.1	5.8	7.7	6.2	1.8	4.0	5.8	0.0	0.0
H2 Avian Hemagglutinin	3.9	3.9	7.5	4.9	2.4	2.8	7.9	8.3	2.8	5.9	8.6	6.9	2.2	3.7	3.7	5.9	7.5	2.0	3.7	5.7	0.0	0.0
Glycoprotein	6.2	5.7	3.6	3.3	5.1	5.7	4.4	6.2	5.1	4.5	6.2	4.9	1.5	1.3	9.5	10.0	14.1	0.3	2.3	3.9	0.0	0.0
Fusion protein	8.7	5.2	7.8	3.7	3.0	3.7	5.6	6.7	0.6	7.6	8.2	4.6	1.9	2.4	3.0	7.1	7.6	0.6	3.2	9.1	0.0	0.0
Nucleoprotein	9.1	6.6	2.9	4.0	0.9	4.0	6.0	8.0	1.4	6.0	10.0	4.6	4.0	3.1	3.7	8.6	5.4	0.6	2.9	8.3	0.0	0.0
Glycoprotein [Avian metapneumovirus type C]	7.0	5.8	6.3	3.9	1.7	4.4	6.8	5.1	1.0	2.6	2.7	8.7	1.2	0.7	7.7	7.2	22.2	0.7	2.2	1.9	0.0	0.0
Fusion protein [Avian metapneumovirus type C	8.4	4.1	5.6	4.5	2.6	2.6	5.8	7.6	0.6	6.3	9.1	6.5	2.2	3.4	3.7	7.4	6.1	0.7	2.6	10.1	0.0	0.0
Haemagglutinin	4.3	3.7	8.2	4.9	2.7	3.4	5.8	5.5	3.0	7.6	7.9	7.3	1.8	3.0	5.8	7.9	5.8	1.8	4.0	5.5	0.0	0.0
Influenza Virus Haemagglutinin	4.3	3.7	8.9	4.3	2.8	3.7	5.8	5.5	3.1	7.4	7.7	6.7	1.8	3.1	5.5	8.0	6.4	1.8	4.0	5.5	0.0	0.0

A=Alanine, Arginine=R, Asparagini=N, Aspartic acid=D, cysteine=C, Glutamic acid=E, Glutamin=Q, Glycine=G, Histidine=H, Isoleucine=I, Leucine=L, Lysine=K, Methionine=M, Phenylalanine=F, Proline=P, Serine=S, Theonine=T, Tryptophan=W, Tyrosine=Y, Valine=V, Selenocystein=U, Pyrrolysine=O

ISSN 2348-1218 (print)
International Journal of Interdisciplinary Research and Innovations ISSN 2348-1226 (online)
Vol. 5, Issue 1, pp: (19-24), Month: January - March 2017, Available at: www.researchpublish.com

Table 4: Prediction of secondary structure of newcastle protein

Protein	Alpha Helix (%)	Beta Turn (%)	Random Coil (%)	Extended Strand (%)
H10 Avian Haemmaglutinin	20.44	12.26	43.40	23.90
H5 Haemagglutinin	13.85	10.77	49.58	25.85
H2 Avian Hemagglutinin	31.43	9.82	35.76	22.99
attachment glycoprotein	18.77	4.11	58.10	19.02
fusion protein	37.36	11.15	28.44	23.05
Nucleoprotein	40.92	10.74	27.37	20.97
attachment glycoprotein [Avian metapneumovirus type C]	7.18	4.96	69.57	18.29
fusion protein [Avian metapneumovirus type C	37.80	10.80	26.82	24.58
Haemagglutinin	20.12	9.76	43.60	26.52
Influenza Virus Haemagglutinin	19.94	8.59	45.71	25.77

Parameters:
Window Width: 17
Similarity Threshold: 8
Number of States: 4.

4. CONCLUSION

This finding observed that out of the ten proteins selected for these study eight proteins shows stability in regard to instability index and extinction coefficient. All the proteins have high amount of aromatic compound which make them stable. This is an indication of resistant to mutation. All the protein indicates thermal instability. All the protein showed high percent composition of serine and threonine amino acid which are quite common in protein functional centre and form hydrogen bond. This typing tool is relevant for subsequent dry and wet laboratory experiement. A further study on pharmacogenetic and mutagenesis regarding this study is needed.

REFERENCES

[1] Aini, I. (1990). Indigenous chicken production in South-East Asia. World. Poult. Sci. J. 46:51-57.

[2] Alexander, D.J. (2003). Newcastle disease, other avian paramyxoviruses, and pneumovirus infection. In: Saif YM et al., (eds). Diseases of poultry. Iowa State University Press, Ames, USA, pp. 64-87.

[3] Banu, N.A., Islam, M.S., Chowdhury, M.M.H. and Islam, M.A. (2009). Determination of immune response of Newcastle disease virus vaccines in layer chickens. J. Bangladesh Agric. Univ. 7(2):329-334.

[4] Barnes, M. R. and Russell, R. B. (1999). A lipid-binding domain in Wnt: a case of mistaken identity? *Curr Biol* 9(19):717–719.

[5] Cai, C.Z., Han, L.Y., Ji, Z.L., Chen, X. and Chen, Y.Z. (2003). SVM-Prot: web-based support vector machine software for functional classification of a protein from its primary sequence. *Nucl. Acids Res*., 31 (13): 3692-3697.

[6] Fennema. (2008). *Food Chemistry 3rd Ed*. CRC Press. pp. 327–8.

[7] Floudas, C.A. (2007). Computational methods in protein structure prediction. Biotechnology and Bioengineering, 97 (2): 207-213.

[8] Gasteiger, E. (2005). (In) John M Walker (edition): The Proteomics Protocols Handbook Humana Press 571-607.

[9] Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D. and Bairoch, A. (2003). ExPASy—the proteomics server for in-depth protein knowledge and analysis. *NucleicAcids Res.* 31: 3784–3788.

[10] Gill, S. C. and Von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* 182: 319–326.

[11] Guruprasad, K., Reddy, B.V. and Pandit, M.W. (1990).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.* 2:155–61.

[12] Ikai, A.J. (1980).Thermostability and aliphatic index of globular proteins. *J Biochem* 88: 1895-1898.\

[13] IUPAC, (1997) *Compendium of Chemical Terminology*, 2nd ed. (the "Gold Book")

[14] Kelley, L.A. and Sternberg, M.J.E. (2009). Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols*,4 (3): 363–371.

[15] Martin, P.A.J. (1992). The Epidemiology of Newcastle disease in village chickens. In: Spradbrow PB (ed) Newcastle disease in village chickens, control with thermostable oral vaccines : Proceeding No. 39 Australian Centre for International Agricultural Research (ACIAR), Caberra, pp. 40-45

[16] Mayo, M.A. (2002). A summary of the changes recently approved by ICTV. Arch. Virol. 147:1655-1656.

[17] Munduganore, D. S., Mundaganore, Y. D. and Ashokan, K. V. (2012). Sequence Analysis of Protein in Pest Des Petits. *International Journal of Food, Agriculture and Veterinary Sciences ISSN: 2277-209X.*

[18] Park, H., Hwang, K.Y., Oh, K.H., Kim, Y.H., Lee, J.Y. and Kim, K.(2008). Discovery of novel alpha-glucosidase inhibitors based on the virtual screening with the homology-modeled protein structure.*Bioorg. Med. Chem*., 16: 284–292.

[19] Rava, P. and Hussain, M.M. (2007). Acquisition of triacylglycerol transfer activity by microsomal triglyceride transfer protein during evolution. *Biochemistry*, 46: 12263–12274.

[20] Wass, M.N., Kelley, L.A. and Sternberg, M.J. (2010). 3DLigandSite: predicting ligand-binding sites using similar structures. *Nucleic Acids Research*, 38: 469-473.

[21] Watson, J.D., Laskowski, R.A. and Thornton, J.M. (2005).Predicting protein function fromsequence and structural data.*Curr.Opin.Struct. Biol*., 15: 275–284.