

Insilico Comparative Modelling and Structure Prediction of Nucleocapsid Protein of Porcine Epidemic Diarrhoea Virus (PEDV) in Pigs

Dhananjay Kashyap¹ and Santosh Kumar^{2*}

P.G. Student, Department of Life Sciences, ITM University, Gwalior, M.P., India¹

Asst. Professor, Department of Life Sciences, ITM University, Gwalior, M.P., India²

ABSTRACT: Porcine epidemic Diarrhoea Virus (PEDV) is a highly contagious pathogen that can cause severe diseases in pigs. It was first known as a devastating enteric disease in feeder and fattening pigs, resembling Transmissible Gastroenteritis in pigs in the United Kingdom, in 1971. It is affiliate to the Coronaviridae family; PEDV infects mainly the epithelial cells of the porcine intestine, leading to acute diarrhoea, vomiting and dehydration. The high morbidity and mortality is caused by PEDV in the newborn piglet (up to two weeks old) among which mortality ranges between 50 -100%. The PEDV, Coronavirus Nucleocapsid (N) protein, plays important roles in the process of virus replication and cellular infection. Virus infection and transfection showed that Nucleocapsid protein was predominately localized in the cytoplasm but also present in nucleolus. Like others RNA viruses, PEDV has high evolutionary rate and is prone to genetic mutations. There are two different types of PEDV that is Type I which only affects growing pigs and Type II is seen in low number. Type II affects all ages, including suckling pigs and mature sows and Type I is a more virulent strain that caused the death of over 1 million pigs in the USA in 2013-2014. Treatment is symptomatic and aims to prevent dehydration in young pigs, using products such as electrolyte and energy supplements. Biosecurity methods such as adequate quarantine, isolation of cases, and disinfection help prevent entry or spread of the disease in the herd. PED is common in the USA, Canada, China, Korea, Japan, Thailand and Vietnam, as well as European countries. Physicochemical characterization of protein isolated from pigs responsible for Porcine Epidemic Diarrhoea disease with 441 amino acids was done to interpret properties like pI, EC, AI, GRAVY and instability index. Insilico comparative modelling was performed to generate good quality models. The assessment of generated three dimensional structure against structure verification tools PROCHECK showed that model generated by Swiss Model was more acceptable to that by Geno3D. The predicted model can be used in structure based drug designing and vaccine development.

KEYWORDS: Porcine Epidemic Diarrhoea Virus (PEDV), Nucleocapsid Protein, PROCHECK, GENO3D

I.INTRODUCTION

Porcine Epidemic Diarrhoea Virus (PEDV) is a Virus that cause Enteric Disease in Swine i.e. Pigs. PEDV is an Alpha corona virus which is a member of Coronaviridae family; this virus is responsible for this deadly disease [1, 9]. PEDV is a major gastrointestinal disease mainly affecting suckling piglets that causes a huge loss in the industrial economics [1]. For the first time PEDV emerged from Europe during the 1970s [2]. PEDV can infect pigs of any age that is from Neonates to Sows but the severity of PED in pigs depends on their ages. PEDV infected neonatal pigs results in death from watery diarrhoea, dehydration, vomiting, anorexia and weight loss [1]. Some microorganisms maintain the normal function of the intestinal mucosal barrier and helps to immune system development. In a study, researchers stated that over 1,000,000 piglets have died from PEDV infection, with a death rate of 80-100% [3, 11]. In the Suckling piglets population this virus causes high mortality [2].

PEDV transmission pathways have been identified, the major mode of transmission is through the faecal-oral routes by direct and indirect contact with the infected piglet or it can spread through Aerosol. Transmission can also occur by contaminated equipment, vehicles used for transport of animals. The pathological changes include the infection of Porcine-enterocytes, blunted intestinal villi, acute mucosal damage and reduced intestinal wall thickening [2, 3].

The PEDV is a 28 kb single stranded, positive sense RNA genome. The genome contains Six known open reading frames (ORFs), such as non-structural protein genes are ORF1a/b and four structural protein genes [Spike (S) , Membrane (M), Envelop (E) and Nucleocapsid (N), there is also an accessory gene present ORF3 [4, 6, 7].

The Corona virus family includes four genera: Alpha corona virus (CoV)-Transmissible Gastroenteritis Virus (TGEV), Beta corona virus (CoV)-Mouse Hepatitis virus (MHV) and Severe Acute Respiratory Syndrome virus (SARS CoV), Gamma corona virus (CoV)- Infectious Bronchitis virus (IBV) and Delta corona virus (CoV) [5, 6].

The PEDV corona virus Nucleocapsid (N) protein plays an important role in Virus replication, RNA synthesis, Phosphorylation and cellular infection. This virus predominates in the Cytoplasm and also found in nucleolus in Vero E6 cells [6]. PEDV mRNA transcribed from the genome produces viral proteins such as spike [S, 180–220 kDa], membrane [M, 27–32 kDa], envelope [E, ~8.8 kDa] and nucleoprotein [N, 55–58 kDa] [4, 6]. N is a multifunctional protein involved in biological processes related to the survival of PEDV. This N protein is localised with major nucleolar proteins including fibrillarin, nucleolin and nucleophosmin [6].

The PEDV entry is mediated by structural type S protein. PEDV replicates exclusively in the cytoplasm. Beside this the N protein has been shown to be involved in PEDV pathogenesis and host cell manipulation. CoV N proteins interact with M proteins for viral assembly and localize with replicase components for viral replication. PEDV N has been shown to inhibit interferon beta (IFN Beta) production and interferon-stimulating gene expression [10].

A huge number of computational tools available makes *Insilico* approaches less time consuming and money saving than that of experimental methods. In this paper the main aim of the study is to do *Insilico* analysis and 3D structure prediction of nucleocapsid protein of PEDV. The structure of this protein has not been described yet. To recognize its various structural features a good quality 3-dimensional model was constructed.

II. MATERIALS AND METHODS

1. Retrieval of target sequence

The amino acid sequence of the nucleocapsid protein of PEDV was obtained from the sequence database of NCBI (<https://www.ncbi.nlm.nih.gov/protein/AVU05351.1/>). It was ascertained that the three-dimensional structure of the protein was not available in Protein Data Bank [12]. Hence the present work for developing the 3D model of the nucleocapsid protein from PEDV was undertaken. The protein is 441 amino acids in length.

2. Physico-chemical characterization

The values of theoretical isoelectric point (pI), molecular weight, total number of positive and negative residues, extinction coefficient [14], instability index [15], aliphatic index [16] and grand average hydropathy (GRAVY) [17] were computed. For physico-chemical characterization, using the ExPASy's ProtParam server [13] (<http://us.expasy.org/tools/protparam.html>). The results were shown in Table 1. Atomic Composition of nucleocapsid protein is shown in Table 2.

3. Secondary structure prediction

Secondary structure has been predicted using PHYRE 2 (<http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index>) software where the FASTA format of the sequence was given as input. SOPMA [12] was employed for calculating the secondary structural features of the protein sequence considered for this study. The results were presented in Table 3.

4. Model building and quality assessment

The modeling of the three dimensional structure of the protein was done using two homology modeling programs, Geno 3D [18] and SwissModel [19] The overall stereochemical property of the protein was assessed by Ramchandran plot analysis [20]. The evaluation of models obtained from the two software tools was performed by using PROCHECK [21]. The models were further checked with WHAT IF [22] for standard bond length and bond angle determination. The results of PROCHECK were shown in Table 4.

III. RESULTS AND DISCUSSION

In the present study, physico-chemical parameters computed using ExPASy's ProtParam tool was represented in Table 1. A protein having instability index smaller than 40 is predicted as stable, on the other hand a value above 40 predicts that the protein may be unstable [15]. The Instability index of 41.14 indicates the unstable nature of nucleocapsid protein. The high extinction coefficient (47900) indicates presence of high concentration of Cys, Trp and Tyr. Aliphatic index (59.71) of query protein suggests that the protein may be stable for a wide temperature range. The energy of the model -4955.78 (kcal/mol) further suggests that the structure is stable. The Atomic composition of PED virus nucleocapsid protein is Carbon (2103), Hydrogen (3371), Nitrogen (671), Oxygen (677) and Sulphur (4) which is computed by ExPASy's ProtParam shown in Table 2. The low value (-1.082) of Grand Average hydropathicity (GRAVY) indicates the possibility of better interaction with water.

SOPMA predicts amino acids correctly to describe secondary structure prediction [23]. The results of SOPMA (shown in table 3) show higher number of random coils in comparison to other secondary structure elements (alpha helix, extended strand and beta turns) and graphs are plotted in Figure 1. Secondary structure and disorder prediction made using phyre 2 software shown in Figure 2 represents that 70.1% of residues were found in the core right handed alpha helices (A), beta sheets (B) and left-handed alpha helix (L) region. 24.3% of residues were found in the allowed right-handed alpha helix (a), beta sheets (b) and left-handed alpha helices regions. The 3.7% of the residues were found in the generously allowed alpha helices (~a), beta sheets (~b), left handed alpha helices (~l) and epsilon (~p) regions. The

1.9% of the residues was found to be localized at the disallowed regions (Figure 3). Above results indicate to a good quality of predicted model.

The homology modelling of the protein was done using two programs Geno3d and SWISS MODEL. The results obtained from these programs were compared (Table-4). This comparison suggests that the model generated by SWISS MODEL is much better than the model generated by Geno 3d. Finally chosen model was visualized by Rasmol (Figure 4).

S. No.	Property	Value
1.	Number of Amino Acid	441
2.	Molecular Weight	49015.23
3.	Theoretical pI	9.90
4.	Total Number of Negatively Charged Residue (Arg + Glu)	72
5.	Total Number of Positively Charged Residue (Arg + Lys)	50
6.	Extinction coefficient	47900
7.	Instability Index	41.14
8.	Aliphatic Index	59.71
9.	Grand Average of Hydropathicity	-1.082

Table 1-Physicochemical parameters computed using Expasy's ProtParam tool

ATOMIC COMPOSITION

S. No.	Atoms	Number of atoms
1.	Carbon (C)	2103
2.	Hydrogen (H)	3371
3.	Nitrogen (N)	671
4.	Oxygen (O)	677
5.	Sulphur (S)	4

Table 2- Atomic composition computed by ExPASy Protparam

SOPMA RESULTS

S. No.	Property	Value (%)
1.	Aplha Helix	21.32
2.	3 ₁₀ Helix	0.00
3.	Pi Helix	0.00
4.	Beta Bridge	0.00
5.	Extended Strand	16.10
6.	Beta Turn	7.26
7.	Bend Region	0.00
8.	Random Coil	55.33
9.	Ambiguous States	0.00
10.	Other States	0.00

Table 3 – Calculated Secondary Structure elements by using SOPMA

Ramchandran Plot-

Software	Parameter	Value (%)
Geno 3D	Residue in most favoured region	70.1
	Residue in additional allowed region	24.3
	Residue in generously allowed region	3.7
	Residue in Disallowed region	1.9
SWISS MODEL	Residue in most favoured region	81.05
	Residue in additional allowed region	14.3
	Residue in generously allowed region	3.7
	Residue in Disallowed region	0.95

Table 4- Comparative study of Ramchandran plot of Geno3D and SWISS MODEL

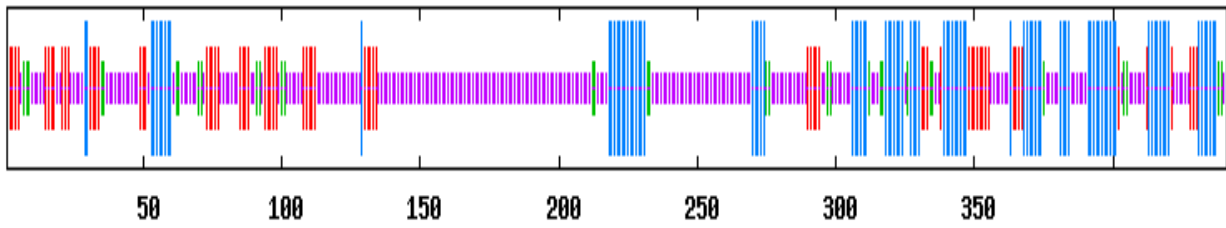


Figure 1- Secondary structure prediction by consensus prediction from multiple alignments

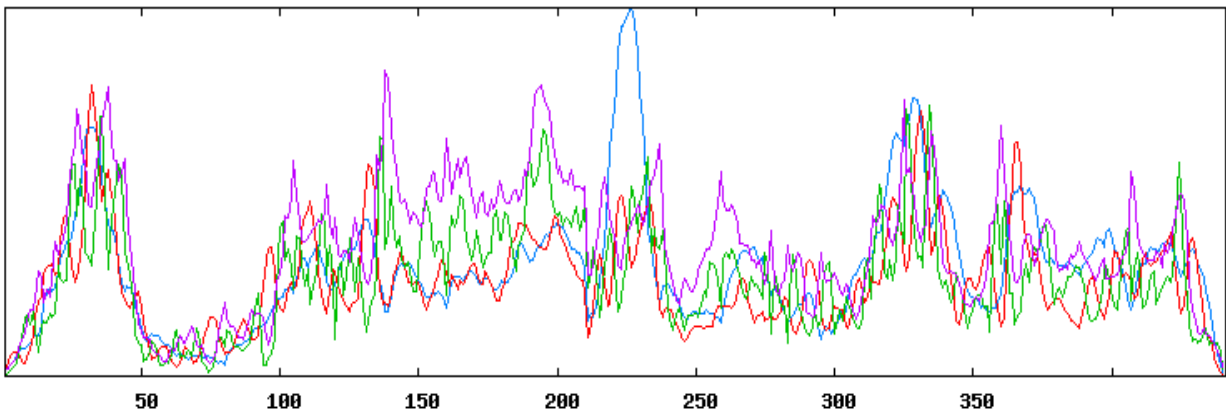


Figure 2: Graphical representation of consensus prediction of secondary structure

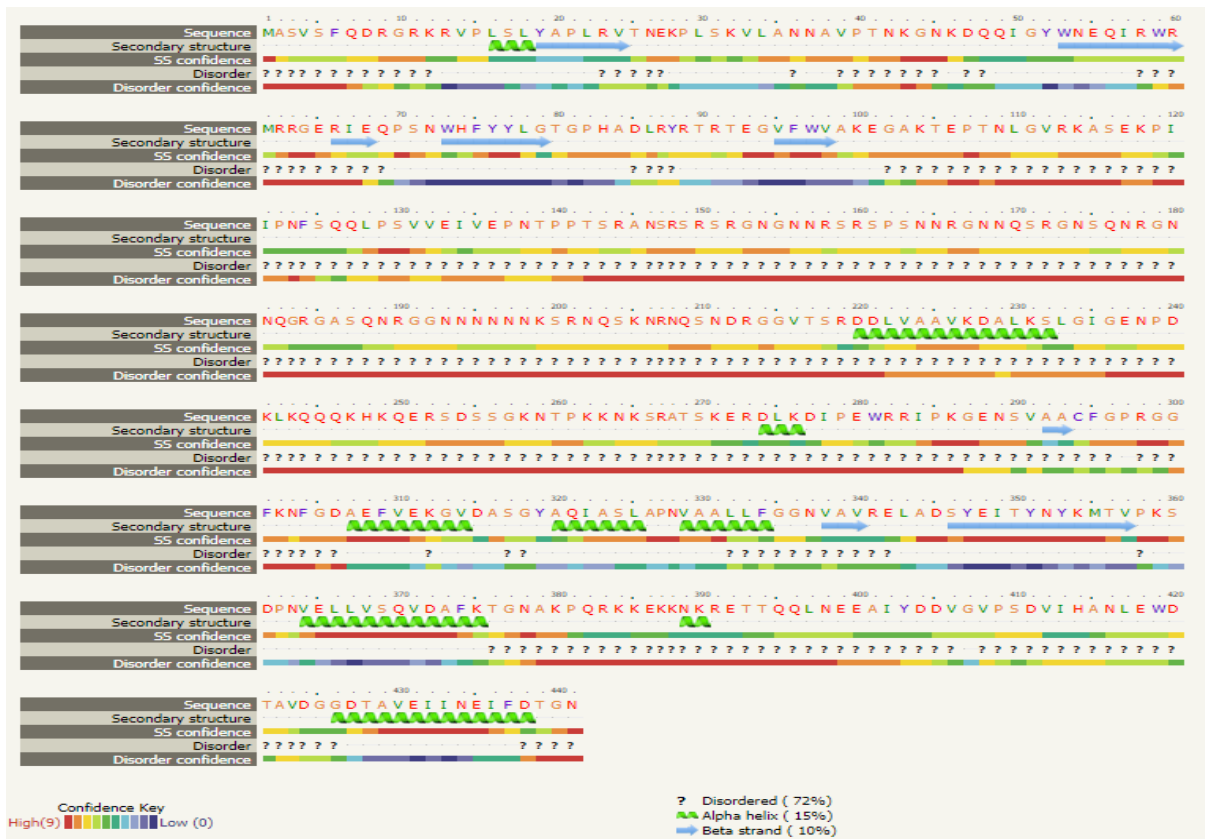


Figure 3- PHYRE 2 Secondary Structure prediction

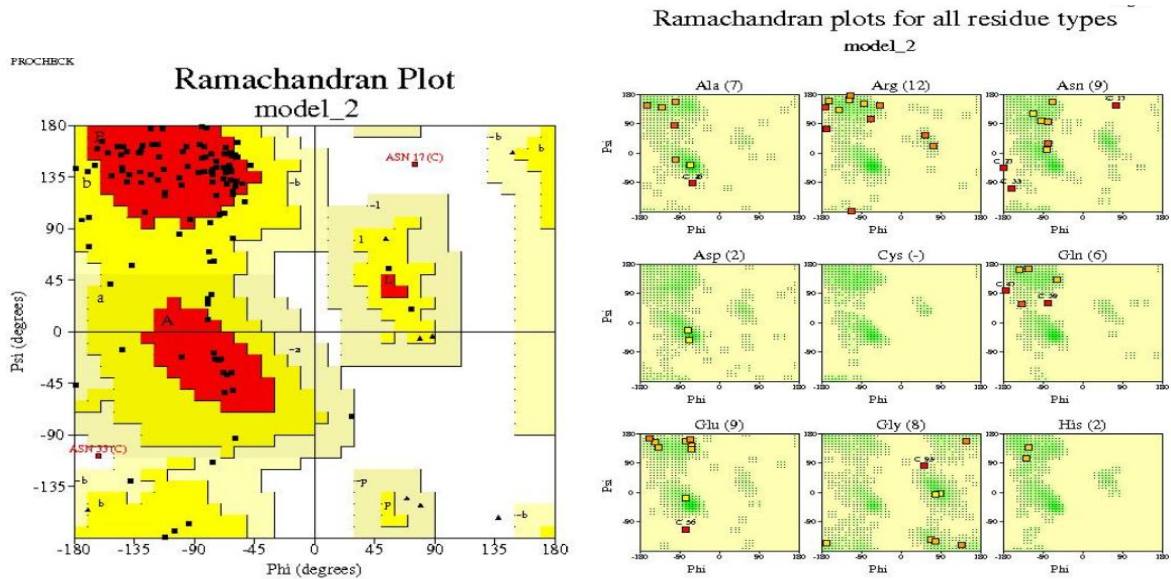


Figure 4: Ramachandran plot of Matrix protein of Porcine Epidemic Diarrhoea Virus generated using PROCHECK software

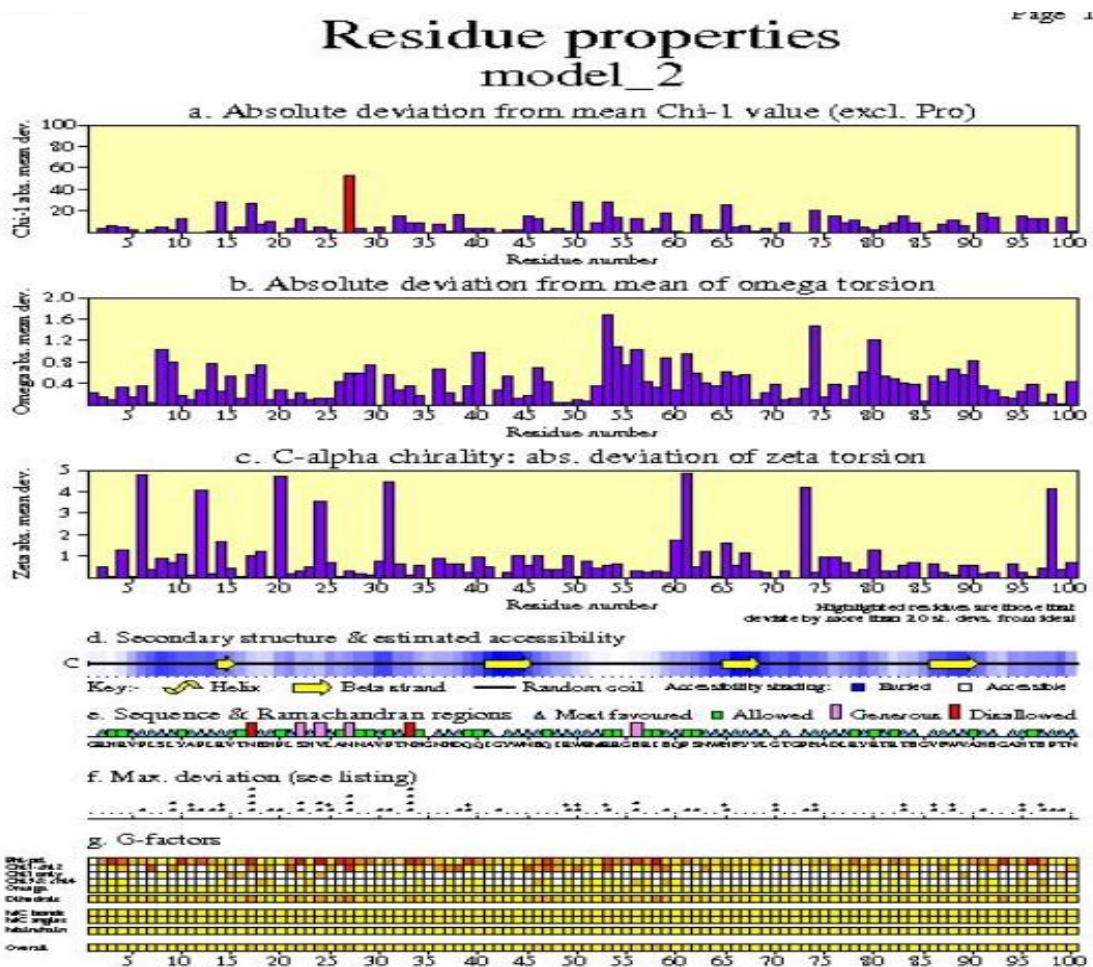


Figure 6: Residual properties and their deviations

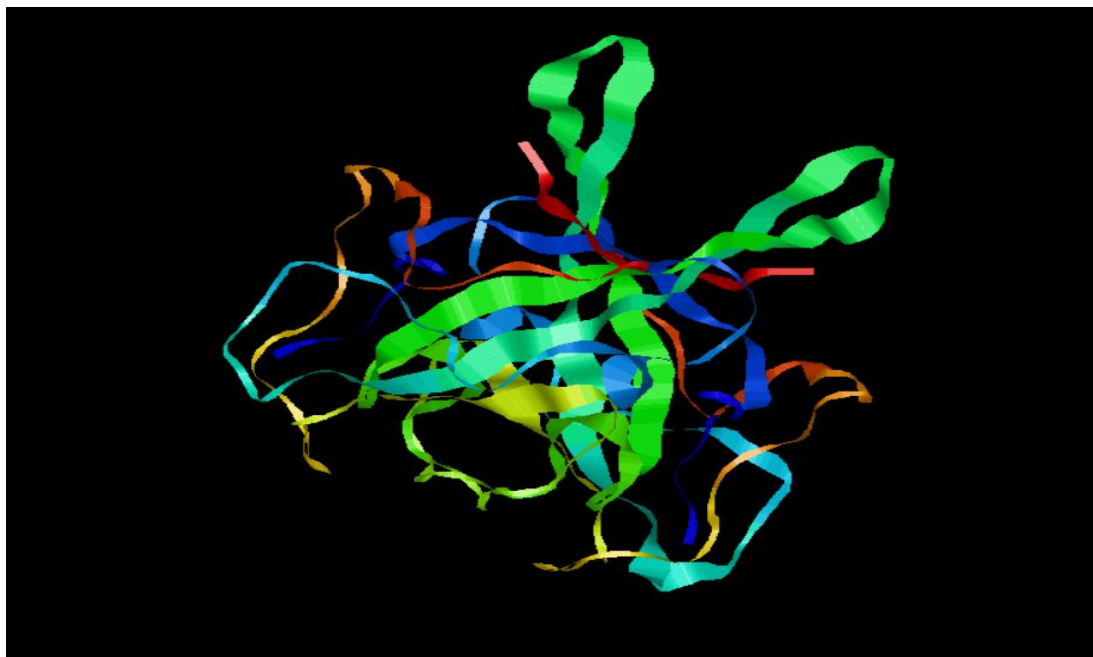


Figure 7: Model Structure of nucleocapsid Protein of Porcine Epidemic Diarrhoea Virus

IV.CONCLUSION

On the basis of various structural and physiochemical parameters assessment, it can be concluded that the predicted three dimensional structure of nucleocapsid protein of Porcine Epidemic Disease Virus is stable. The Ramchandran plot data showed that nearly 72% of the residues for each protein lies in the core region. The maximum percentage of the residues lying in the core region represents the best stereo chemical quality of the protein structure which is further supported by its nearly 81% residues lying in the most favoured and allowed regions. Only 0.95 % residues lie in the disallowed regions which generally involve steric hindrance between the side chain C-beta methylene group and main chain atoms. Since the percentage residues lying in the disallowed region is much lower than its presence in core and allowed regions the structures generated show best stereochemical quality and stability. The three dimensional structures generated using SWISS MODEL software and validated by PROCHECK software showed the best identity, lowest energy in Kcal/mol and root mean square deviation (RMSD) were considered as the best optimized and superimposed model. Since no effective therapeutic or vaccine is available for Bloom syndrome. Structural information of this model can be effectively used and can be further implemented in future drug designing

V.ACKNOWLEDGEMENTS

We would like to acknowledge ITM University, Gwalior for its timely support and also the groups who supported us directly or indirectly in making this research a success. We would also like to thank the softwares without whose support this work would have not been completed.

REFERENCES

- 1- Zhen Tan 1,2, Wanting Dong 1, Yaqun Ding 1, Xiangdong Ding 1, Qin Zhang 1,3 and Li Jiang “Porcine Epidemic Diarrhea Altered Colonic Microbiota Communities in Suckling Piglets” <https://www.mdpi.com/2073-4425/11/1/44>
- 2- Gallien et al. Vet Res (2018) 49:7 “Evidence of porcine epidemic diarrhea virus (PEDV) shedding in semen from infected specific pathogen-freeboars” <https://veterinaryresearch.biomedcentral.com/track/pdf/10.1186/s13567-018-0505-2>
- 3- Yixuan Hou and Qihong Wang “Emerging Highly Virulent Porcine Epidemic Diarrhea Virus: Molecular Mechanisms of Attenuation and Rational Design of Live Attenuated Vaccines” <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6862049/>
- 4- Dang Wang,a,b Liurong Fang,a,b Yanling Shi,a,b Huan Zhang,a,b Li Gao,a,b Guiqing Peng,a,b Huanchun Chen,a,b Kui Li,c Shaobo Xiaoa,b “Porcine Epidemic Diarrhea Virus 3C-Like Protease Regulates Its Interferon Antagonism by Cleaving NEMO”

- <https://jvi.asm.org/content/90/4/2090>
- 5- Da Shi †, Maojie Lv †, Jianfei Chen, Hongyan Shi, Sha Zhang, Xin Zhang and Li Feng * “Molecular Characterizations of Subcellular Localization Signals in the Nucleocapsid Protein of Porcine Epidemic Diarrhea Virus”
<https://www.mdpi.com/1999-4915/6/3/1253>
 - 6- Liang Li,a Fang Fu,a Shanshan Guo,a Hongfeng Wang,b Xijun He,a Mei Xue,a Lingdan Yin,a Li Feng,a Pinghuang Liua “Porcine Intestinal Enteroids: a New Model for Studying Enteric Coronavirus Porcine Epidemic Diarrhea Virus Infection and the Host Innate Response”
<https://jvi.asm.org/content/93/5/e01682-18>
 - 7- Ying SHAN1, Zi-qi LIU1, Guo-wei LI1, Cong CHEN1, Hao LUO1, Ya-jie LIU1, Xun-hui ZHUO2, Xing-fen SHI3, Wei-huan FANG1, Xiao-liang LI† “Nucleocapsid protein from porcine epidemic diarrhea virus isolates can antagonize interferon-λ production by blocking the nuclear factor-κB nuclear translocation”
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6052364/>
 - 8- Benjamas Liwnaree, Jaraspim Narkpuk, Suttipun Sungsuwan, Anan JongkaewwattanaID, Peera Jaru-Ampornpan “Growth enhancement of porcine epidemic diarrhea virus (PEDV) in Vero E6 cells expressing PEDV nucleocapsid protein”
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0212632>
 - 9- Zhen Ding, Liurong Fang, Huiyuan Jing, Songlin Zeng, Dang Wang, Lianzeng Liu, Huan Zhang, Rui Luo, Huanchun Chen, Shaobo Xiao “Porcine Epidemic Diarrhea Virus Nucleocapsid Protein Antagonizes Beta Interferon Production by Sequestering the Interaction between IRF3 and TBK1”
<https://jvi.asm.org/content/88/16/8936>
 - 10- Peera Jaru-Ampornpan,a Juggragarn Jengarn,a,b Asawin Wanitchang,a Anan Jongkaewwattana “Porcine Epidemic Diarrhea Virus 3C-Like Protease-Mediated Nucleocapsid Processing: Possible Link to Viral Cell Culture Adaptability” <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5215342/>
 - 11- Daesub Song1*, Hyoungjoon Moon2*, Bokyu Kang2 “Porcine epidemic diarrhea: a review of current epidemiology and available vaccines”
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4524901/>
 - 12- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN and Bourne PE, The Protein Data Bank. *Nucleic Acids Res*, 28 (1): 235-42, 2000.
 - 13- Gasteiger E (2005) Protein Identification and Analysis Tools on the ExPASy Server. In: John M. Walker, The Proteomics Protocols Handbook, Humana Press 571-607.
 - 14- Gill SC, Von Hippel PH (1989) Extinction coefficient. *Anal Biochem* 182: 319-328
 - 15- Guruprasad K, Reddy BVP, Pandit MW (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. *Prot Eng* 4: 155-16
 - 16- Ikai AJ (1980) Thermo stability and aliphatic index of globular proteins. *J Biochem* 88: 1895-1898
 - 17- Kyte J, Doolittle RF (1982) A simple method for displaying the hydrophobic character of a protein. *J Mol Biol* 157: 105- 132
 - 18- Combet C, Jambon M, Deleage G, Geourjon C (2002) Geno3D: Automatic comparative molecular modelling of protein. *Bioinformatics* 18: 213-214
 - 19- Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22: 195-201
 - 20- Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. *J Mol Biol* 7: 95-99
 - 21- Laskowski RA, Rullmann JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 8: 477-486
 - 22- Vriend G (1990) WHAT IF: A molecular modeling and drug design program. *J Mol Graph* 8: 52-56
 - 23- Geourjon C, Deléage G (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput Appl Biosci* 11: 681-684.