Article

Decoding the characteristics of ORF6 encoded protein of Norway rat Hepatitis E Virus using bioinformatics approach

Zoya Shafat[1], **Anwar Ahmed**[2], **Mohammad K. Parvez**[3], **Shama Parveen**[1]

[1]Centre for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, New Delhi, India

[2]Centre of Excellence in Biotechnology Research, College of Science, King Saud University, Riyadh, Saudi Arabia

[3]Department of Pharmacognosy, College of Pharmacy, King Saud University, Riyadh, Saudi Arabia

E-mail: zoya179695@st.jmi.ac.in, anahmed@ksu.edu.sa, mohkhalid@ksu.edu.sa, sparveen2@jmi.ac.in

Received 25 October 2021; *Accepted 29 November 2021*; *Published 1 March 2022*

Abstract

Hepatitis E virus (HEV) of the family *Hepeviridae*, is a major causative agent of acute hepatitis in developing countries. The Norway rat HEV genome is organized into six open reading frames (ORFs), i.e., ORF1, ORF2, ORF3, ORF4, ORF5 and ORF6. The additional reading frame encoded protein ORF6 is attributed to life cycle of rat HEV. As ORF6 protein's remains to be explored in terms of its structural and functional implications, the following study was conceptualized to explore the prospective role of this additional genomic component of rat HEV. The detailed computational investigation was carried out for the ORF6 protein to elucidate its physiochemical properties, primary structure, secondary structure, tertiary structure and post-translational modifications, motif prediction and other functional characteristics. The *in silico* analysis revealed ORF6 protein as unstable, highly thermostable, hydrophobic and basic in nature. The amino acid compositional analysis revealed higher abundance of Leu, Arg, Ile and Pro amino acids in the polypeptide chain of ORF6 protein. The secondary structural analysis revealed all the three major elements, i.e., α-helices, β-strands and coils. The generated 3D structural model of the ORF6 protein through homology modeling algorithm revealed mixed α/β structural fold of the ORF6 protein with abundance of coils. Additionally, the structural models revealed the presence of clefts and a tunnel. The identified binding functions and the presence of several clefts suggested the commitment of ORF6 protein towards interaction with other ligand molecules. This theoretical study will facilitate towards deciphering the role of unexplored ORF6 encoded protein, thereby providing better understanding towards the pathogenesis of Norway rat HEVs.

Keywords rat HEV; open reading frame 6 (ORF6); physicochemical parameters; structural analysis; homology modeling; motif prediction; molecular function; biological function.

Network Biology
ISSN 2220-8879
URL: http://www.iaees.org/publications/journals/nb/online-version.asp
RSS: http://www.iaees.org/publications/journals/nb/rss.xml
E-mail: networkbiology@iaees.org
Editor-in-Chief: WenJun Zhang
Publisher: International Academy of Ecology and Environmental Sciences

1 Introduction

Hepatitis E virus (HEV) is the major aetiological agent of Hepatitis E, also called enteric hepatitis (enteric means related to the intestines) infection (Kumar et al., 2013). Worldwide, about 20 million HEV infections and 3.3 million symptomatic hepatitis E cases occur annually which results in 44,000 deaths (Khuroo and Khuroo, 2016). HEV belongs to the family *Hepeviridae* and genus *Orthohepevirus* (Takahashi et al., 2010). The HEV genome is a single, positive-sense RNA (7.2 kb in length), which is flanked with short 5′ and 3' non-coding regions (NCR) (Tam et al., 1991). The HEV genome is categorized into three open reading frames (ORFs): ORF1, ORF2 and ORF3. The ORF1, ORF2 and ORF3 encode the non-structural polyprotein (pORF1), capsid protein (pORF2) and the pleotropic protein (pORF3) respectively (Kenney and Meng, 2019). The HEV ORF1 is categorized into seven domains: methyltransferase (MTase/MeT), Y (undefined), papain-like cysteine protease (PCP), hypervariable region/proline-rich hinge (HVR/PPR), X (macro), helicase (Hel/NTPase), and RNA-dependent RNA polymerase (RdRp) (Ansari et al., 2000; Ropp et al., 2000; Suppiah et al., 2011; Parvez, 2017). Rodent also plays a role in the epidemiology of hepatitis E. In the year 2009, two complete nucleotide sequences were analyzed from Norway rats in Germany which suggested a completely separate genotype for these HEV strains (Johne et al., 2010). These nucleotide sequences had high divergence to other HEV strains, i.e., HEV G1, HEV G2, HEV G3, HEV G4 and avian HEV (Johne et al., 2010). It was predicted through software that the genome in these rat HEV sequences was organized into a total of six reading frames (ORF1, ORF2, ORF3, ORF4, ORF5 and ORF6). i.e., rat HEV genome consisted of three additional ORFs (ORF4, ORF5 and ORF6). It was also identified that unlike typical HEV genomic organization, the ORFs ORF1 and ORF33 do not overlap in these two rat HEVs. Three additional putative ORFs of 280 - 600 nt that overlap with ORFs 1 or 2 were predicted for each rat HEV genome (Johne et al., 2010).

Recent studies have elucidated the characteristics of some of the less understood ORF encoded proteins using computational approaches to delineate their role in the pathogenesis of HEV (Shafat et al., 2021a, 2021b, 2021c). Though ORF6 is attributed to Norway rat HEVs genomic organization, however, this additional frame encoded protein has not been explored in terms of its structure and function. Thus, we conceptualized this present study to determine the characteristics of the ORF6 protein of rat HEVs. We conducted computational analyses to provide insights into the structural and functional characteristics of this potential region.

2 Material and Methods

2.1 Sequence retrieval

The rat HEV ORF6 amino acid sequence (Accession number: GU345042) was retrieved from the GenBank database obtained from NCBI (National Center for Biotechnology Information).

2.2 Physicochemical properties analysis

The amino acid sequences of the ORF6 protein in FASTA format was used as query in for the determination of physiochemical parameters. The various physical and chemical parameters of the retrieved sequences were computed using ProtParam (Expasy), a web-based server (Gasteiger et al., 2005). The ProtParam tool employed various parameters such as, instability index (II – protein stability) (Guruprasad et al., 1990), aliphatic index (AI – relative volume occupied by protein's aliphatic side chains) (Ikai A, 1980), extinction coefficients (EC – protein-protein/protein-ligand interactions quantitative study) (Gill and Hippel, 1989), Grand Average of Hydropathicity (GRAVY - sum of all hydropathicity values divided by number of residues in a sequence) (Kyte and Doolittle 1982), theoretical pI, half-life (Gonda et al., 1989), and number of positive and negative residues.

2.3 Primary structural analysis

The primary structure of the ORF6 protein in terms of the percentage composition of amino acids was

computed using the ProtParam (Expasy) tool and PSIPRED (PSIPRED Workbench (ucl.ac.uk).

2.4 Secondary structural analysis

The secondary structure analysis of the ORF6 protein was conducted using two different webservers, i.e., self-optimized prediction method with alignment (SOPMA) software (npsa-prabi.ibcp.fr) and PSIPRED (PSIPRED Workbench (ucl.ac.uk).

2.5 Tertiary structural analysis

The tertiary structure of the target protein was modeled using the online program Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2). The generated ORF6 protein 3D model was validated using Ramachandran plot analysis (PROCHECK) (http://nihserver.mbi.ucla.edu/SAVES) for stereo-chemical property.

2.6 Functional analysis

N-linked glycosylation prediction: The N-linked sites for glycosylation were predicted using NetNGlyc 1.0 (http://www.cbs.dtu.dk/services/NetNGlyc/) server**,** provided by Centre for Biological Sequence Analysis, Technical University of Denmark (CBS DTU). ***O-linked glycosylation prediction:*** The O-linked sites for glycosylation were predicted using NetOGlyc 4.0 (http://www.cbs.dtu.dk/services/NetOGlyc/) server, provided by Centre for Biological Sequence Analysis, Technical University of Denmark (CBS DTU). ***Phosphorylation prediction:*** The phosphorylation sites were predicted using NetPhos3.1 (NetPhos - 3.1 - Services - DTU Health Tech) server, provided by Centre for Biological Sequence Analysis, Technical University of Denmark (CBS DTU). For phosphorylation studies, we performed both generic and kinase specific predictions. ***Motifs prediction*****:** ANTHEPROT v.6.9.3 predicted phosphorylation and other modified sites in the ORF6 protein. ***Peptide signal detection:*** Location of signal peptide cleavage in the ORF6 protein was predicted using Signal P-4.1 (SignalP - 5.0 - Services - DTU Health Tech). ***Molecular function, biological process and cellular component prediction:*** I-TASSER COFACTOR algorithm was used to identify the cellular localization of the ORF6 protein.

3 Results

3.1 Analysis of physicochemical properties

The various physiochemical parameters of the ORF6 protein are summarized in Table 1.

Table 1 Physicochemical parameters of the ORF6 protein of rat HEV.

Physicochemical Properties	**ORF6**
Number of amino acids	91
Molecular weight	10257.35
Theoretical pI	10.77
Total number of negatively charged residues (Asp + Glu)	2
Total number of positively charged residues (Arg + Lys)	10
Formula	$C_{469}H_{751}N_{131}O_{113}S_7$
Total number of atoms	1471
Extinction coefficient (assuming all Cys pairs residues form cystines)	19730
Extinction coefficient (assuming all Cys pairs residues are reduced)	19480
Estimated half-life	30 hours (mammalian reticulocytes, in vitro) > 20 hours (yeast, in

	vivo) > 10 hours (*Escherichia coli*, in vivo)
Instability index	67.79
Aliphatic index	112.53
Grand average of hydropathicity (GRAVY)	0.443

Physicochemical analysis showed that this polypeptide is of 91 amino acids, with an isoelectric point (pI) of 10.77. The instability index was 67.79, which classified the protein as unstable (as >40 value implies unstable protein). The positive aliphatic index value (112.53) suggested it as a highly thermostable protein. The positive GRAVY value of the ORF6 protein was found predicted its hydrophobic nature as the positive score indicates the hydrophobicity. Thus, taken together it can be interpreted that the ORF6 protein was found to be unstable, highly thermostable, hydrophobic and basic in nature.

3.2 Analysis of primary structure

Proteins differ from one another in their structure, primarily in their sequence of amino acids. The linear sequence of the amino acid polypeptide chain refers to its primary structure. The amino acid composition of ORF6 protein is summarized in Table 2 (Fig. 1).

Table 2 Amino acid composition of the ORF6 protein of rat HEV.

Amino acid	ORF6
Ala (A)	6.6
Arg (R)	9.9
Asn (N)	1.1
Asp (D)	1.1
Cys (C)	4.4
Gln (Q)	2.2
Glu (E)	1.1
Gly (G)	6.6
His (H)	3.3
Ile (I)	8.8
Leu (L)	14.3
Lys (K)	1.1
Met (M)	3.3
Phe (F)	3.3
Pro (P)	8.8
Ser (S)	6.6
Thr (T)	6.6
Trp (W)	3.3
Tyr (Y)	2.2
Val (V)	5.5
Pyl (O)	0
Sec (U)	0

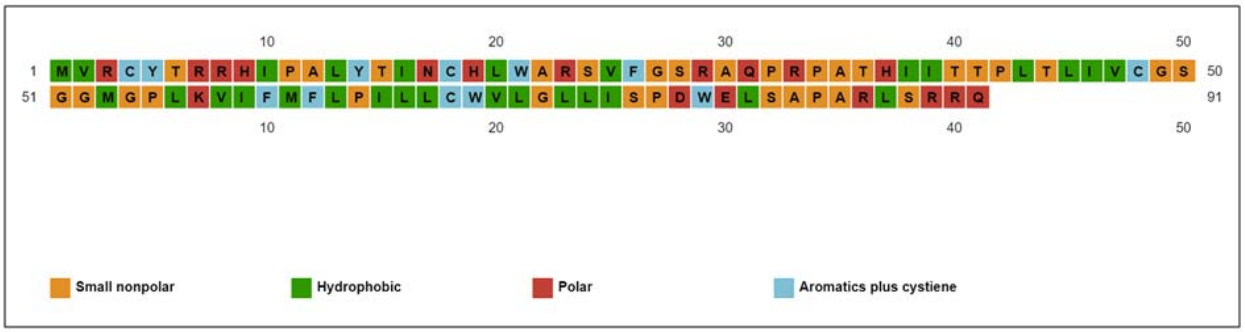

Fig. 1 Representation of amino acid composition of ORF6 protein. The analysis was conducted using PSIPRED.

Leu was observed as the top contributing amino acid to the ORF6's polypeptide chain. The top five amino acids that contributed to the polypeptide chain of ORF6 were included Leu, Arg, Ile, Pro and Ser/Thr/Gly (Fig. 1).

3.3 Analysis of secondary structure

The default parameters (similarity threshold: 8; window width: 17) were considered by SOPMA for the secondary structure prediction with >70% prediction accuracy, utilizing 511 proteins (sub-database) and 15 aligned proteins. The predicted elements of secondary structure in the ORF6 protein are mentioned in Table 3.

Table 3 Secondary structure elements of the ORF6 protein of rat HEV by SOPMA.

S. No.	Secondary structure elements	Values (%)
1	Alpha helix	34.07
2	3_{10}helix	0.00
3	Pi helix	0.00
4	Beta bridge	0.00
5	Extended strand	24.18
6	Beta turn	8.79
7	Bend region	0.00
8	Random coil	32.97
9	Ambiguous states	0.00
10	Other states	0.00

*The values are represented as percentages.

Thus, taken together, SOPMA predicted that the ORF6 protein consisted of all the three major elements of secondary structure, i.e., alpha helix (α), beta strand (β) and random coil. The predicted secondary structure element by PSIPRED is shown in Fig. 2.

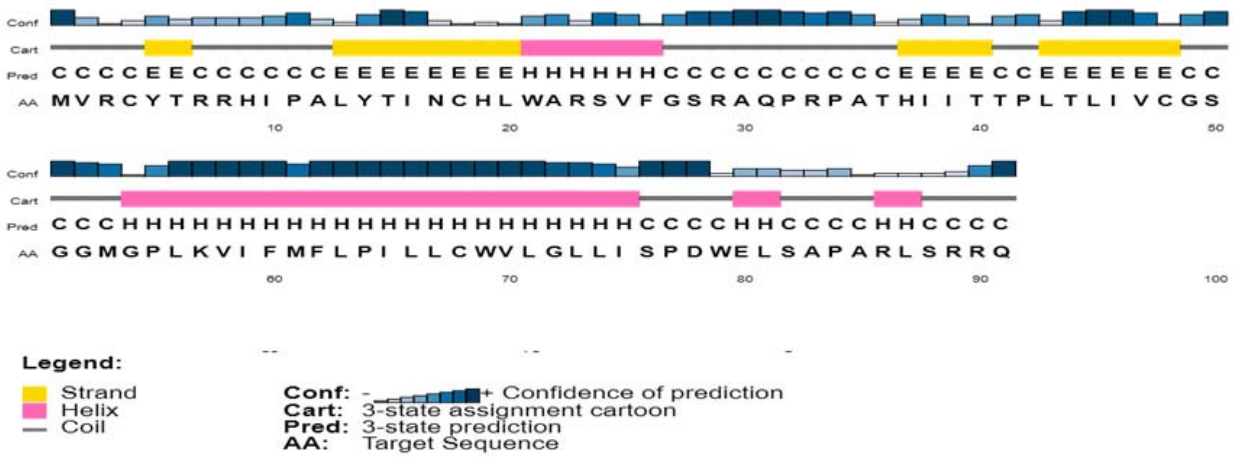

Fig. 2 Secondary structure elements of ORF6 protein of rat HEV. The analysis was conducted using PSIPRED.

3.4 Analysis of tertiary structure

The amino acids structural diversity plays a vital role in the formation of protein self-assembly. The three-dimensional spatial arrangement of amino acid residues in a protein is known as the tertiary structure. The secondary structure elements (helices and strands) are combined in different ways to form three-dimensional structures of a protein. To perform structure-based drug-designing, it is quite essential to build a reliable model. The generated 3D tertiary structure of the ORF6 protein (via Phyre2) was analyzed by visualization through homology modeling approach (Table 4) (Fig. 3).

Table 4 Properties of the modeled 3D structure generated through Phyre2.

Model dimensions (Å)	Template	Secondary structure and disorder prediction
X: 32.914	C2jr3A	Disordered (22%)
Y: 44.986		Alpha helix (30%)
Z: 44.792		Beta strand (33%)
		TM helix (34%)

Fig. 3 Tertiary structure of the ORF6 protein of rat HEV. The analysis was conducted using Phyre2 webserver.

Further, the obtained 3D ORF6 model (generated through Phyre2) was evaluated using PROCHECK (Fig. 4).

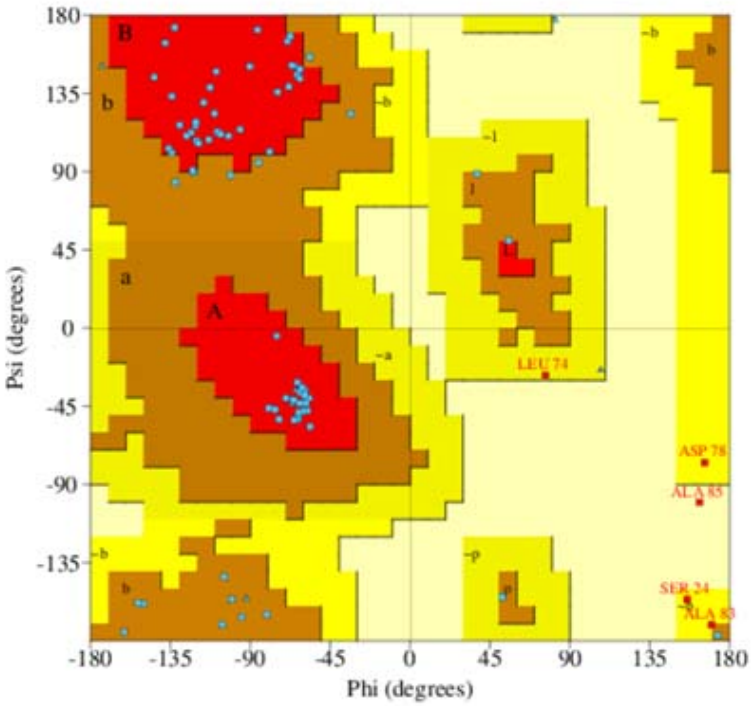

Fig. 4 Ramachandran plot of the ORF6 protein of rat HEV showing the favoured region**s.** The analysis was conducted using PROCHECK.

The overall protein's stereochemical quality, amino acids present in the allowed, disallowed region and the G-factor was evaluated by Ramachandran map (Table 5).

Table 5 Statistics for the obtained 3D ORF6 model.

PDBsum analysis	
Clefts	10
Pores	-
Tunnels	1
PROCHECK analysis	
Ramachandran Plot statistics	
Most Favoured Regions	70.7%
G-Factors	
Overall average	-0.43

Based on an analysis of **118** structures of resolution of at least **2.0** Angstroms and *R*-factor no greater than **20.0** a good quality model would be expected to have over **90%** in the most favoured regions [A,B,L].
***G-factors** provide a measure of how **unusual**, or out-of-the-ordinary, a property is.
Values below -0.5* - unusual
Values below **-1.0**** - highly unusual

Further, the obtained 3D structural model was also identified with the presence of several clefts and a tunnel (Fig. 5).

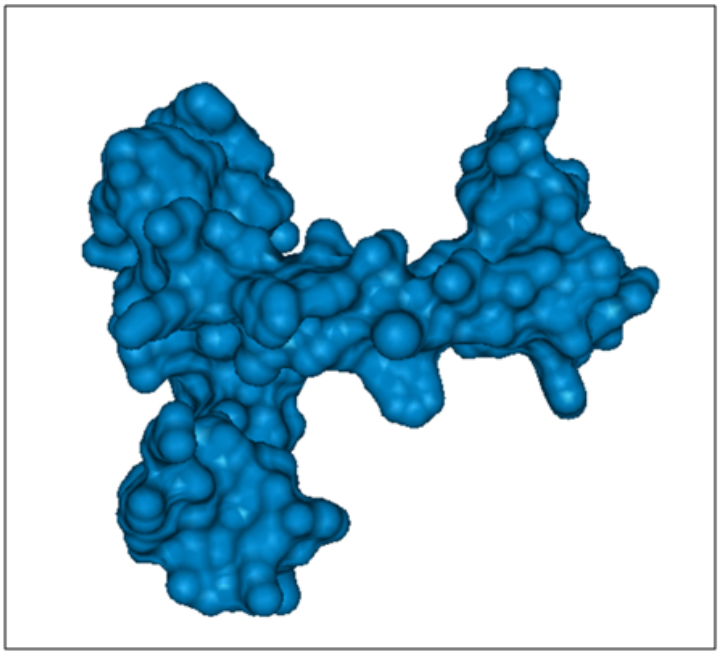

Fig. 5 Surface representation of the modeled 3D structure of the ORF6 protein of rat HEV.

3.5 Analysis of functional characteristics

None of the N-linked and O-linked possible sites for glycosylation were identified in the ORF6 protein. However, several phosphorylation sites including 3 Ser and 4 Thr residues were identified in the ORF6 protein using NetPhos3.1 server (Fig. 6).

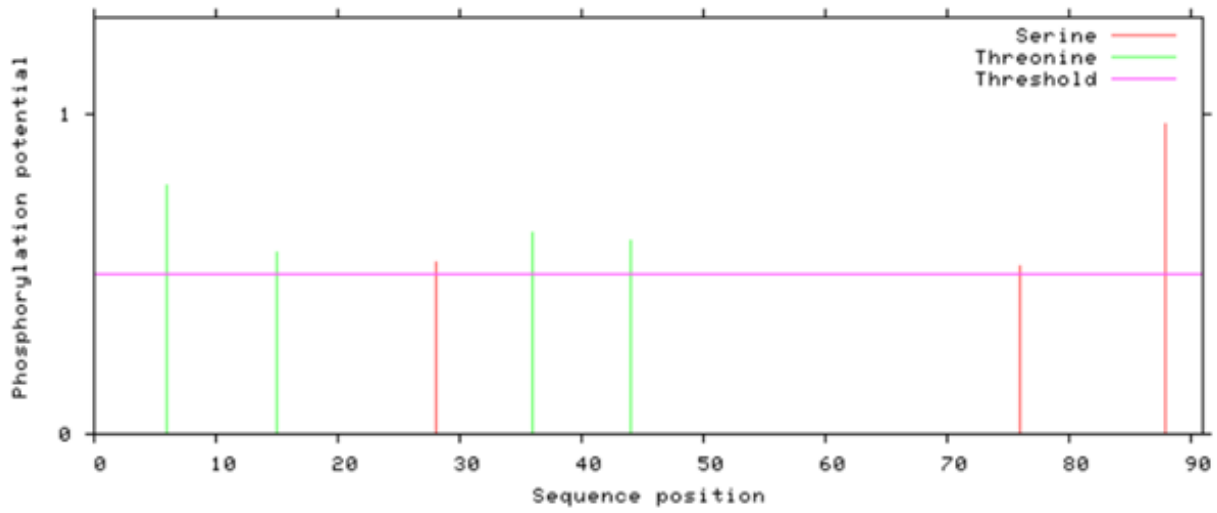

Fig. 6 Predicted phosphorylation sites using NetPhos3.1.

Additionally, it was revealed through ANTHEPROT that the ORF6 protein contained two protein kinase C phosphorylation sites. The potential cleavage site for signal peptide were found to be absent in the amino acid sequence (Fig. 7).

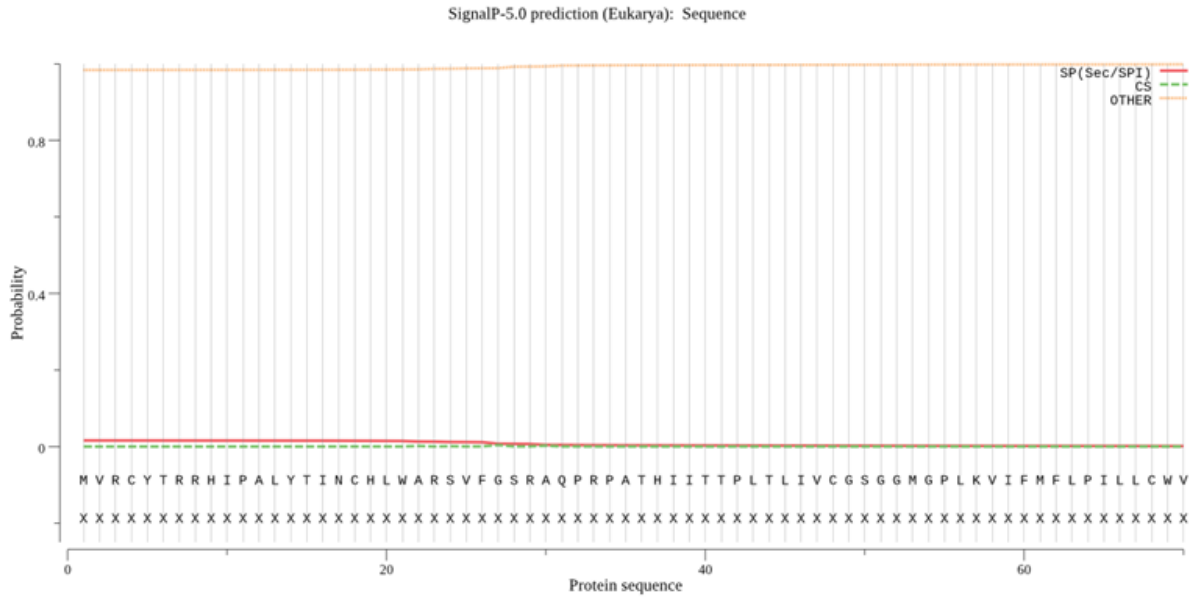

Fig. 7 SignalP-5.0 prediction. Signal peptide likelihood was absent.

The consensus GO annotations associated to the 3D ORF6 model (generated through I-TASSER) through CO-FACTOR algorithm is mentioned in Table 6.

Table 6 Predicted consensus GO terms for rat HEV ORF6 model.

Molecular function	
GO:0008800~beta-lactamase activity	Catalysis of the reaction: a beta-lactam + H2O = a substituted beta-amino acid.
GO:0050660~flavin adenine dinucleotide binding	Binding to FAD, flavin-adenine dinucleotide, the coenzyme or the prosthetic group of various flavoprotein oxidoreductase enzymes, in either the oxidized form, FAD, or the reduced form, FADH2.
GO:0018659~4-hydroxybenzoate 3-monooxygenase activity	Catalysis of the reaction: 4-hydroxybenzoate + NADPH + H+ + O2 = protocatechuate + NADP+ + H2O.
GO:0004222~ metalloendopeptidase activity	Catalysis of the hydrolysis of internal, alpha-peptide bonds in a polypeptide chain by a mechanism in which water acts as a nucleophile, one or two metal ions hold the water molecule in place, and charged amino acid side chains are ligands for the metal ions.
GO:0046872~ metal ion binding	Binding to a metal ion.
Biological Process	

GO:0017001~ antibiotic catabolic process	The chemical reactions and pathways resulting in the breakdown of antibiotic, a substance produced by or derived from certain fungi, bacteria, and other organisms, that can destroy or inhibit the growth of other microorganisms.
GO:0043639~ benzoate catabolic process	The chemical reactions and pathways resulting in the breakdown of benzoate, the anion of benzoic acid (benzenecarboxylic acid), a fungistatic compound widely used as a food preservative; it is conjugated to glycine in the liver and excreted as hippuric acid.
GO:0046677~ response to antibiotic	Any process that results in a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of an antibiotic stimulus.
GO:0019439~aromatic compound catabolic process	The chemical reactions and pathways resulting in the breakdown of aromatic compounds, any substance containing an aromatic carbon ring.
GO:0006508~proteolysis	The hydrolysis of proteins into smaller polypeptides and/or amino acids by cleavage of their peptide bonds.
Cellular Component	
GO:0030288~outer membrane-bounded periplasmic space	The region between the inner (cytoplasmic or plasma) membrane and outer membrane of organisms with two membranes such as Gram negative bacteria.
GO:0016020~membrane	A lipid bilayer along with all the proteins and protein complexes embedded in it an attached to it.

The molecular functions included flavin adenine dinucleotide binding (GO:0050660), metal ion binding (GO:0046872), beta-lactamase activity activity (GO:0008800) and 4-hydroxybenzoate 3-monooxygenase activity (GO:0018659). Thus, binding interactions and catalytic activities were the major functional roles that were attributed to ORF6 protein. Furthermore, the predicted different biological processes revealed the significant roles associated with the ORF6 protein in rat HEV.

4 Discussion

The Norway rat HEV genome comprises six ORFs (ORF1, ORF2, ORF3, ORF4, ORF5 and ORF6). Although ORF6 is attributed to genomic component of HEV, its functional implications remains unexplored (Johne et al., 2010). Thus, in the presented study, we examined the structural and functional properties of the ORF6 encoded protein through evaluating its physicochemical parameters, different structural levels (primary structure, secondary structure, tertiary structure), functional characteristics, such as post-translational modifications, motif prediction and gene ontology analysis, using the combination of different computer tools. The availability of the Norway rat HEV sequences, consisting additional ORFs (ORF4, ORF5 and ORF6) in

GenBank database encouraged us to inspect the characteristics of the ORF6 protein.

The physiochemical parameters are imperative in decoding the protein's characteristics, thus were analyzed. Initially, the various physico-chemical properties of the ORF6 protein were computed using the amino acid sequence through ProtParam (ExPasy server) (Gasteiger et al., 2005). The parameters included molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY) (Gasteiger et al., 2005). The half-life is the amount of time which is taken by the protein to disappear its half of the amount after its synthesis in the cell (Bachmair et al., 1986; Gonda et al., 1989; Tobias et al., 1991; Ciechanover and Schwartz, 1989; Varshavsky, 1997). In this study, the half-life of all the proteins was 30 h. Instability index is another crucial factor which estimates the protein's stability in a test tube. A protein having instability index value is above the value of 40 is predicted as unstable, while a value smaller than 40 predicts it as stable. Our higher instability index (>60) value indicated the unstable nature of the ORF6 protein (Guruprasad et al., 1990). The aliphatic index is another important factor that governs the protein nature (Ikai, 1980). It is defined as the relative volume occupied by aliphatic side chains (Ala, Val, Iso and Leu). The positive aliphatic index score is directly proportional to the increased thermostability of the globular proteins, i.e., proteins having higher aliphatic indices are comparatively more thermally stable in comparison to proteins having lesser aliphatic indices (Ikai, 1980). Thus, high aliphatic index value indicated high thermostability possessed by the ORF6 protein due to the presence of some amino acids having aliphatic side chains (Ala, Val, Ile and Leu). Additionally, GRAVY is considered as an important factor for protein in determining its physiochemical properties (Kyte and Doolittle, 1982). The positive value of GRAVY suggested the hydrophobic nature of the ORF6 protein as the value between - 0.310 and - 0.514 and lower values have been shown to have good interaction with water indicating the hydrophilic nature of the given protein (Kyte and Doolittle, 1982).

The amino acid composition revealed the higher abundance of the amino acids Leu, Arg, Ile, Pro and Ser/Thr/Gly in the ORF6 proteins. Leu is categorized into the group "regulatory" as this group consists of eight most potent amino acids, such as Tyr, Phe, Gln, Pro, His, Trp, and Met (Mortimore and Pösö, 1987; Garlick, 2005). Charged amino acids (such as, Arg) contribute to the process of ligand binding (Claverie and Notredame, 2007). Pro has important functional and structural implications as it performs important role, such as molecular recognition, intracellular signaling and essential signaling cascades (Kay et al., 2000). The amino acid Ser is generally classified as a nutritionally nonessential (dispensable), but plays a crucial role in different cellular processes (Kalhan and Hanson, 2012). Thus, it could be interpreted that our initial compositional analysis in terms of significant amino acid contribution to the ORF6's polypeptide chain signifies its role in various regulatory functions.

The ORF6 protein 3D structure modeled by Phyre2 also showed both α and β content with subsequently higher percentage of coils. Thus, our tertiary structural analysis was in agreement with the secondary structural analysis. Thus, it could be interpreted that the ORF6 protein is a mixed α/β structural-fold with predominance of coils. Additionally, identification of clefts, tunnels and pores accessible to ligand molecules is essential in the context of structure-based drug design process (Mbarek et al., 2019; Marques et al., 2017). Thus, the modeled 3D structure of the OR6 protein was scrutinized to reveal the presence of binding sites. Interestingly, the PDBsum analysis of the modeled ORF6 protein structure revealed several clefts and a tunnel. Clefts present on protein's surface are sizeable depressions and are important in determining the protein interaction with other molecules as they possess tendency to be enzyme active sites (Coleman and Sharp, 2006). The size of clefts is considered as primary factors in governing the interaction between the receptor protein with the target molecules (Coleman and Sharp, 2006). Tunnels are defined as access paths which

connects the interior of the protein molecule to the surrounding environment. Furthermore, tunnels influence the reactivity of the protein and determine the interaction nature and intensity (Brezovsky et al., 2018). Thus, the presence of clefts and tunnel also strengthens our analysis, suggesting the commitment of ORF6 protein towards interaction with the target molecules.

The ORF6 protein model was predicted with motifs which included modified sites such as phosphorylation. Protein phosphorylation constitutes an essential mechanism for the proper establishment of an infection cycle in several intracellular pathogens (Marks, 1996; Zor et al., 2002). Phosphorylation is required for protein folding, signal transduction, intracellular localization PPIs, transcription regulation, cell cycle progression, survival and apoptosis (Vihinen et al., 2001; Li et al., 1990; Keck et al., 2015). Thus, taken together, it can be interpreted from our findings that the ORF6 protein performs crucial functions by interacting with the other viral and host components, thus signifying its essentiality in rat HEV pathogenesis.

5 Conclusions

The ORF6 is attributed as an additional genomic component to Norway rat HEV. Due to lack of information on the additional reading frame encoded protein, this study was conceptualized to determine the functional and structural characteristics of the ORF6 protein of the rat HEV. The *in silico* analysis revealed that the ORF6 protein was highly unstable, thermostable, hydrophobic and basic in nature. The structural analysis revealed mixed α/β structural fold of the ORF6 protein with predominance of coils. Identification of the complete structure of the ORF6 protein will provide insights into its functional role in the viral pathogenesis. Our theoretical data could assist in further research to explore the biology of the ORF6 protein of rat HEV. Additionally, detailed experimental confirmations of these analyses are envisaged towards a better understanding of the HEV biology.

Acknowledgement

The authors would like acknowledge Maulana Azad National Fellowship (MANF) supported by the Government of India.

References

Ansari IH, Nanda SK, Durgapal H, Agrawal S, Mohanty SK, Gupta D, Jameel S, Panda SK. 2000. Cloning, sequencing, and expression of the hepatitis E virus (HEV) nonstructural open reading frame 1 (ORF1). Journal of Medical Virology, 60(3): 275-283

Bachmair A, Finley D, Varshavsky A. 1986. In vivo half-life of a protein is a function of its amino-terminal residue. Science, 234: 179-186

Brezovsky J, Kozlikova B, Damborsky J. 2018. Computational analysis of protein tunnels and channels. In :Protein Engineering. 25-42, Humana Press, New York, USA

Ciechanover A, Schwartz AL. 1989. How are substrates recognized by the ubiquitin-mediated proteolytic system? Trends in Biochemical Sciences, 14: 483-488

Coleman RG, Sharp KA. 2006. Travel depth, a new shape descriptor for macromolecules: application to ligand binding. Journal of Molecular Biology, 362(3): 441-458

Garlick PJ. 2005. The role of leucine in the regulation of protein metabolism. The Journal of Nutrition, 135(6): 1553S-1556S

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. 2005. Protein identification and analysis tools on the ExPASy Server. In: The Proteomics Protocols Handbook. 571-607,

Humana Press, USA

Gill SC, Hippel PHV. 1989. Calculation of protein extinction coefficient from amino acid sequence data. Analytical Biochemistry, 182: 319-326

Gonda DK, Bachmair A, Wunning I, Tobias JW, Lane WS, Varshavsky A. 1989. A Universality and structure of the N-end rule. Journal of Biological Chemistry, 264: 16700-16712

Guruprasad K, Reddy BVB, 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. Protein Engineering, Design and Selection, 4: 155-161

Ikai A. 1980. Thermostability and aliphatic index of globular proteins. The Journal of Biochemistry, 88: 1895-1898

Jean-Michel C, Notredame C. 2007. Bioinformatics for dummies, Wiley Publishing, 111: 07030-5774

Johne R, Heckel G, Plenge-Bönig A, Kindler E, Maresch C, Reetz J, Schielke A, Ulrich RG. 2010. Novel hepatitis E virus genotype in Norway rats, Germany. Emerging infectious diseases, 16(9): 1452

Kalhan SC, Hanson RW. 2012. Resurgence of serine: an often neglected but indispensable amino Acid. J Biological Chemistry, 287(24): 19786-19791

Kay BK, Williamson MP, Sudol M. 2000. The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. The FASEB Journal, 14(2): 231-241

Keck F, Ataey P, Amaya M, Bailey C, Narayanan A. 2015. Phosphorylation of single stranded RNA virus proteins and potential for novel therapeutic strategies. Viruses, 7(10): 5257-5273

Kenney SP, Meng XJ. 2019. Hepatitis E virus genome structure and replication strategy. Cold Spring Harbor perspectives in medicine, 9(1): a031724

Khuroo MS, Khuroo MS. 2016. Hepatitis E: an emerging global disease–from discovery towards control and cure. Journal of Viral Hepatitis, 23(2): 68-79

Kumar S, Subhadra S, Singh B, Panda BK. 2013. Hepatitis E virus: the current scenario. International Journal of Infectious Diseases, 17(4): e228-e233

Kyte J, Doolittle RF. 1982. A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology, 157: 105–132

Li G, La Starza MW, Hardy WR, Strauss JH, Rice CM. 1990. Phosphorylation of Sindbis virus nsP3 in vivo and in vitro. Virology, 179(1): 416-427

Marks F. 1996. Protein Phosphorylation. VCH

Marques SM, Daniel L, Buryska T, Prokop Z, Brezovsky J, Damborsky J. 2017. Enzyme tunnels and gates as relevant targets in drug design. Medicinal Research Reviews, 37(5): 1095-1139

Mbarek A, Moussa G, Leblond Chain J. 2019. Pharmaceutical applications of molecular tweezers, clefts and clips. Molecules, 24(9): 1803

Mortimore GE, Pösö AR. 1987. Intracellular protein catabolism and its control during nutrient deprivation and supply. Annual Review of Nutrition, 7(1): 539-568

Parvez MK. 2017. The hepatitis E virus nonstructural polyprotein. Future Microbiology, 12(10): 915-924

Ropp SL, Tam AW, Beames B, Purdy M, Frey TK. 2000. Expression of the hepatitis E virus ORF1. Archives of Virology, 145(7): 1321-1337

Shafat Z, Ahmed A, Parvez MK, Parveen S. 2021. Role of "dual-personality" fragments in HEV adaptation—analysis of Y-domain region. Journal of Genetic Engineering and Biotechnology, 19: 154

Shafat Z, Ahmed A, Parvez MK, Parveen S. 2021. Role of ORF4 in Hepatitis E virus regulation: analysis of intrinsically disordered regions. Journal of Proteins and Proteomics

Shafat Z, Ahmed A, Parvez MK, Parveen S. 2021. Sequence to structure analysis of the ORF4 protein from

Hepatitis E virus. Bioinformation, 17(9): 818-828

Suppiah S, Zhou Y, Frey TK. 2011. Lack of processing of the expressed ORF1 gene product of hepatitis E virus. Virology Journal, 8(1): 1-5

Takahashi M, Tanaka T, Takahashi H, Hoshino Y, Nagashima S, Mizuo H, Yazaki Y, Takagi T, Azuma M, Kusano E, Isoda N. 2010. Hepatitis E Virus (HEV) strains in serum samples can replicate efficiently in cultured cells despite the coexistence of HEV antibodies: characterization of HEV virions in blood circulation. Journal of Clinical Microbiology, 48(4): 1112-1125

Tam AW, Smith MM, Guerra ME, Huang CC, Bradley DW, Fry KE, Reyes GR. 1991. Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length viral genome. Virology, 185(1): 120-131

Tobias JW, Shrader TE, Rocap G, Varshavsky A. 1991. The N-end rule in bacteria. Science 254: 1374-1377

Varshavsky A. 1997. The N-end rule pathway of protein degradation. Genes to Cells, 2(1): 13-28

Vihinen H, Ahola T, Tuittila M, Merits A, Kääriäinen L. 2001. Elimination of phosphorylation sites of Semliki Forest virus replicase protein nsP3. Journal of Biological Chemistry, 276(8): 5745-5752

Zor T, Mayr BM, Dyson HJ, Montminy MR, Wright PE. 2002. Roles of phosphorylation and helix propensity in the binding of the KIX domain of CREB-binding protein by constitutive (c-Myb) and inducible (CREB) activators. Journal of Biological Chemistry, 277(44): 42241-42248