



Sequence to structure analysis of the ORF4 protein from Hepatitis E virus

Zoya Shafat¹, Anwar Ahmed², Mohammad K. Parvez³ & Shama Parveen^{1*}

¹Centre for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, New Delhi, India; ²Centre of Excellence in Biotechnology Research, College of Science, King Saud University, Riyadh, Saudi Arabia; ³Department of Pharmacognosy, College of Pharmacy, King Saud University, Riyadh, Saudi Arabia; Corresponding author: Shama Parveen - Email: sparveen2@jmi.ac.in

Author contacts:

Zoya Shafat - zoya179695@st.jmi.ac.in; Anwar Ahmed - anahmed@ksu.edu.sa; Mohammad K. Parvez - mohkhalid@ksu.edu.sa

Received August 14, 2021; Revised September 21, 2021; Accepted September 21, 2021, Published September 30, 2021

DOI: 10.6026/97320630017818

Declaration on Publication Ethics:

The authors state that they adhere with COPE guidelines on publishing ethics as described elsewhere at <https://publicationethics.org/>. The authors also undertake that they are not associated with any other third party (governmental or non-governmental agencies) linking with any form of unethical issues connecting to this publication. The authors also declare that they are not withholding any information that is misleading to the publisher in regard to this article.

Author responsibility:

The authors are responsible for the content of this article. The editorial and the publisher have taken reasonable steps to check the content of the article in accordance to publishing ethics with adequate peer reviews deposited at PUBLONS.

Declaration on official E-mail:

The corresponding author declares that official e-mail from their institution is not available for all authors

Abstract:

Hepatitis E virus (HEV) is the main cause of acute hepatitis worldwide. HEV accounts for up to 30% mortality rate in pregnant women with highest incidences reported for genotype 1 (G1) HEV. The contributing factors in adverse cases during pregnancy in women due to HEV infection is still debated. The mechanism underlying the pathogenesis of viral infection is attributed to different genomic components of HEV, i.e., open reading frames (ORFs): ORF1, ORF2, ORF3 and ORF4. Recently, ORF4 has been discovered in enhancing the replication of GI isolates of HEV through regulation of an IRES-like RNA element. However, its characterization through computational methodologies remains unexplored. In this novel study, we provide comprehensive overview of ORF4 protein's genetic and molecular characteristics through analyzing its sequence and different structural levels. A total of three different datasets (Human, Rat and Ferret) of ORF4 genomes were built and comparatively analyzed. Several non-synonymous mutations in conjunction with higher entropy values were observed in rat and ferret datasets, however, limited variation was observed in human ORF4 genomes. Higher transition to transversion ratio was observed in the ORF4 genomes. Studies have reported the association of intrinsic disordered proteins (IDPs) with drug discovery due to its role in several signaling and regulatory processes through protein-protein interactions (PPIs). As PPIs are potent drug target sources, thus the ORF4 protein was explored by analyzing its polypeptide structure in order to shed light on its intrinsic disorder. Pressures that lead towards preponderance of disordered-promoting amino acid residues shaped the evolution of ORF4. The intrinsic disorder propensity analysis revealed ORF4 protein (Human) as a highly disordered protein (IDP). Predominance of coils and lack of secondary structure further substantiated our findings suggesting its involvement in binding to ligand molecules. Thus, ORF4 contributes to cellular signaling processes through protein-protein interactions, as IDPs are targets for regulation to accelerate the process of drug designing strategies against HEV infections.

Keywords: Hepatitis E virus (HEV), open-reading frame 4 (ORF4), mutational analysis, entropy analysis, gene selection pressure, nucleotide diversity, transition/transversion ratio, structural analysis, intrinsic disordered proteins (IDPs), drug target

Background:

Hepatitis E virus (HEV) is a major causative agent of viral hepatitis transmitted enterically worldwide. HEV is the major aetiological agent of Hepatitis E, also called enteric hepatitis (enteric means related to the intestines) infection [1]. HEV is an *Orthohepevirus* [3], with a single-strand, positive-sense RNA genome of around 7.2 Kb in length and flanked with short 5' and 3' non-coding regions (NCR) [4]. The HEV genome comprises three partially overlapped 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 [5]. Further studies led to the identification of a novel viral protein synthesized from an ORF within ORF1, which was named as ORF4. This newly identified ORF4 was first reported by Nair *et al.* [6] which is exclusive to HEV G1. The indispensability of ORF4 in viral replication has been demonstrated. It has been revealed that ORF4 interacts with multiple viral and host proteins to enhance virus replication [6, 7]. The expression of this ORF4 protein is regulated via an internal ribosome entry site (IRES)-like RNA element that is unregulated via cellular endoplasmic reticulum (ER) stress. ORF4 protein is rapidly turned over within cells as it possesses a proteasomal degradation signal [6]. Additionally, ORF4 has also been recognized in rats and ferrets [8, 9]. Though ORF4 essentiality in G1 viral replication has been determined, its genetic and molecular characteristics remain to be explored.

Thus, in the present study we have analyzed the functional and structural characteristics of the ORF4 proteins by exploiting sequence-based bioinformatics methods. The data analysis of pathogen's genomic sequences has been progressively increased in the past few decades. At present, it is considered as an important approach in the epidemiology of infectious [10]. Availability of large number of complete genomic sequences of HEV on the (NCBI) has been achieved due to incredible effort made by researchers worldwide accumulating HEV data. This available data enabled us to comprehend the molecular basis of the evolution/ genomic variability/molecular biology in ORF4 region of HEV. In this context, a comparative codon-based characterization of the HEV ORF4 was conducted in an attempt to estimate the evolutionary divergence in the ORF4 gene sequences of HEV genome. The findings may contribute towards predicting the signature sequences based on the codon-based model of molecular evolution. Till date, specific treatment against HEV strains has not been discovered. Only Hecolin, a prophylactic vaccine is licensed only in China [11]. Thus, further studies are required for the development of specific drug molecule to treat HEV infections against all strains. Drug-discovery has been associated with intrinsically disordered proteins (IDPs) due to their prime features [12]. IDPs lack well-defined stable structure but are significantly involved in several biological processes, such as various signaling and regulatory processes, in protein-protein interactions (PPIs) [13 - 15]. Usually, IDPs form hub proteins in PPI networks [12]. Studies have reported the close association of various IDPs with several human diseases (tumor, Parkinson disease, Alzheimer disease, diabetes) [16 - 20]. The disease associated IDPs perform crucial roles in the disease through PPI networks [15]. IDPs undergo couple binding and folding end exist as ensembles of interconverting structures [21]. Due to the involvement in numerous PPIs, IDPs are considered as potential drug targets for drug molecules, which are capable of modulating or inhibiting their interactions, thus have opened tremendous potential in the field of drug discovery [22]. Very recently, the indispensability of ORF4 in HEV replication has been demonstrated [8]. In this context, we conducted computational analyses to provide an

insight into the structural characteristics of this potential region. Therefore, the intrinsically disordered regions of the ORF4 proteins of HEV were analyzed. The findings obtained from the present analysis are augmented to envisage our understanding towards the biology of ORF4 protein of HEV.

Materials and methods**Sequence data acquisition**

The HEV ORF4 sequences were obtained from the National Centre for Biotechnology Information (NCBI) GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) public database. The complete detail of the sequences considered for the present analysis is listed in **Table 1**.

Multi-sequence alignment of ORF4 protein genes

The ORF4 sequences considered for the present analyses were categorized into three datasets. Dataset I consisted of study sequences from the host organism Human. Dataset II contained study sequences from the host organism Rat. Dataset III contained study sequences from the host Ferret. The alignment for all three datasets was achieved using Clustal X2 in BioEdit v.7.2 [23].

Mutational analysis of ORF4 protein genes

Bioedit software was used to predict the amino acid substitution in the HEV study sequences encompassing human, rat and ferret

Analysis of entropy of ORF4 protein genes

The Shannon entropy analysis of ORF4 was carried out using BioEdit software [23], for the identification of possible variability/mutability. The entropy of aligned amino acids sequences was calculated at particular codon position to comprehend the variation within these genes.

Selection pressure analysis of ORF4 protein genes

Mutation rates were determined for ORF4 gene sequences using Gene selection pressure. Gene selective pressure was estimated by Tajima test of neutrality implemented in MEGAX software [24]. Positive selection was considered when D value is found to be positive (greater than 0). The test compares the average number of nucleotide differences between pairs of sampled sequences (referred to as pairwise difference-) and the total number of polymorphic sites (segregating site- S) in the sampled DNA sequences. The difference in the expectations for these two variables (which can be positive or negative) defines the Tajima's D test statistic. A positive Tajima's D signifies selection while negative D specifies purifying selection.

Codon degeneracy patterns estimation of ORF4 protein genes

The estimation of different codon values including nucleotide diversity (π), number of segregating sites (S) and transition to transversion ratio (R) in ORF4 gene sequences was undertaken for all the datasets. The analysis was conducted using the MEGAX software [24].

Structural analysis of ORF4 proteins

Recent study on intrinsic disordered proteins (IDPs) revealed that they have the potential to act as drug targets [17 - 20]. Therefore, we evaluated the different structure levels of ORF4 proteins, obtained from different sources, to shed some light on its sequence composition, secondary structure elements, intrinsic disorder content and binding tendency. Thus, a set of different computational prediction methods was exploited to determine the structure of ORF4 protein.

PSIPRED (<http://bioinf.cs.ucl.ac.uk/psipred/>), SOPMA (Self-Optimized Prediction Method with Alignment) (<https://npsa->

prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html) and PONDR-fit (<http://original.disprot.org/pondr-fit.php>), TASSER (Iterative Threading ASSEMBLY Refinement) [25] and PROCHECK [26]. The structural analyses were conducted using four different ORF4 protein sequences, i.e., LC057248 (HEV) KU168733 (Human), JN167538 (Rat) and LC177791 (Ferret).

Table 1: Demographics of HEV ORF4 genomes analyzed in present study

No	Strain	GenBank Accession	Host	Country	Year
1	R63	NC_038504	Rat	Germany	2009
2	31479	KU168736	Human	India	2014
3	31390	KU168735	Human	India	2014
4	29714	KU168734	Human	India	2014
5	27370	KU168733	Human	India	2013
9	HEV324	LC549185	Rat	China	2012
10	HEV577	LC549184	Rat	China	2012
14	Mu09/0434	JN167538	Rat	Germany	2009
15	Mu09/0685	JN167537	Rat	Germany	2009
16	R68	GU345043	Rat	Germany	2009
17	SF4370	LC057247	Ferret	Japan	2013
18	HEV-4342	AB890374	Ferret	Japan	2013
19	HEV-4351	AB890001	Ferret	Japan	2013
20	F63	LC177792	Ferret	USA	2016
21	F61	LC177791	Ferret	USA	2016
22	F60	LC177790	Ferret	USA	2016
23	F54	LC177789	Ferret	USA	2016
24	F45	LC177788	Ferret	USA	2016
25	FRHEV20	JN998607	Ferret	Netherland	2010
26	FRHEV4	JN998606	Ferret	Netherland	2010
27	R63	GU345042	Rat	Germany	2009
28	31865	KU168737	Human	India	2014

Table 2: The specific codon positions along with non-synonymous substitutions in the ORF4 genomes of Hepatitis E viruses

Dataset I		
T(66)M	KU168735/India/2014/Homo sapiens KU168734/India/2013/Homo sapiens KU168737/India/2013/Homo sapiens KU168736/India/2013/Homo sapiens	
Dataset II		
S(2)W	GU345043/Germany/2009	
E(11)G	LC549185/China/2012 LC549184/China/2012 GU345043/Germany/2009	
R(18)Q	LC549185/China/2012/Rattus losea LC549184/China/2012/Rattus norvegicus JN167538/Germany/2009/Rattus norvegicus JN167537/Germany/2009/Rattus norvegicus	
Y(29)F	LC549185/China/2012/Rattus losea LC549184/China/2012/Rattus norvegicus	
P(31)H	JN167538/Germany/2009/Rattus norvegicus	
T(32)S	JN167537/Germany/2009/Rattus norvegicus	
T(35)I	LC549185/China/2012/Rattus losea LC549184/China/2012/Rattus norvegicus JN167538/Germany/2009/Rattus norvegicus JN167537/Germany/2009/Rattus norvegicus	
P(38)L	LC549185/China/2012/Rattus losea LC549184/China/2012/Rattus norvegicus	
L(39)P	LC549185/China/2012/Rattus losea LC549184/China/2012/Rattus norvegicus	
S(40)C	LC549185/China/2012/Rattus losea LC549184/China/2012/Rattus norvegicus	
S(40)F	JN167537/Germany/2009/Rattus norvegicus GU345043/Germany/2009/Rattus norvegicus	
S(44)F	LC549185/China/2012/Rattus losea JN167538/Germany/2009/Rattus norvegicus JN167537/Germany/2009/Rattus norvegicus GU345043/Germany/2009/Rattus norvegicus	
S(45)F	LC549184/China/2012/Rattus norvegicus	
N(53)S	LC549185/China/2012/Rattus losea LC549184/China/2012/Rattus norvegicus JN167538/Germany/2009/Rattus norvegicus JN167537/Germany/2009/Rattus norvegicus	
P(54)L	JN167538/Germany/2009/Rattus norvegicus	
A(64)V	LC549185/China/2012/Rattus losea LC549184/China/2012/Rattus norvegicus GU345043/Germany/2009/Rattus norvegicus	
K(71)R	LC549185/China/2012/Rattus losea	
M(73)T	LC549184/China/2012/Rattus norvegicus JN167537/Germany/2009/Rattus norvegicus LC549185/China/2012/Rattus losea	
G(75)A	LC549184/China/2012/Rattus norvegicus JN167538/Germany/2009/Rattus norvegicus JN167537/Germany/2009/Rattus norvegicus LC549185/China/2012/Rattus losea	
V(76)A	LC549184/China/2012/Rattus losea LC549185/China/2012/Rattus losea LC549184/China/2012/Rattus norvegicus JN167538/Germany/2009/Rattus norvegicus JN167537/Germany/2009/Rattus norvegicus LC549185/China/2012/Rattus losea	
P(80)L	LC549184/China/2012/Rattus losea LC549185/China/2012/Rattus losea LC549184/China/2012/Rattus norvegicus JN167538/Germany/2009/Rattus norvegicus JN167537/Germany/2009/Rattus norvegicus GU345043/Germany/2009/Rattus norvegicus	
L(84)S	LC549185/China/2012/Rattus losea LC549184/China/2012/Rattus norvegicus JN167538/Germany/2009/Rattus norvegicus	
D(89)G	LC549184/China/2012/Rattus norvegicus	
T(94)M	LC549185/China/2012/Rattus losea	
F(98)Y	JN167538/Germany/2009/Rattus norvegicus JN167537/Germany/2009/Rattus norvegicus	
T(99)I	JN167538/Germany/2009/Rattus norvegicus JN167537/Germany/2009/Rattus norvegicus	
F(102)S	LC549185/China/2012/Rattus losea	
F(103)S	LC549184/China/2012/Rattus losea LC549185/China/2012/Rattus losea	
H(104)R	JN167538/Germany/2009/Rattus norvegicus LC549185/China/2012/Rattus losea LC549184/China/2012/Rattus norvegicus JN167538/Germany/2009/Rattus norvegicus JN167537/Germany/2009/Rattus norvegicus	
L(105)Q	LC549185/China/2012/Rattus losea	
T(106)M	JN167538/Germany/2009/Rattus norvegicus	
A(112)V	LC549185/China/2012/Rattus losea LC549184/China/2012/Rattus norvegicus JN167538/Germany/2009/Rattus norvegicus	
I(115)T	LC549185/China/2012/Rattus losea	
A(116)V	LC549185/China/2012/Rattus losea LC549184/China/2012/Rattus norvegicus	
G(118)D	JN167537/Germany/2009/Rattus norvegicus	
A(119)G	LC549185/China/2012/Rattus losea LC549184/China/2012/Rattus norvegicus	
A(119)V	JN167538/Germany/2009/Rattus norvegicus	
G(120)A	LC549185/China/2012/Rattus losea LC549184/China/2012/Rattus norvegicus	
R(121)Q	LC549185/China/2012/Rattus losea LC549184/China/2012/Rattus norvegicus JN167538/Germany/2009/Rattus norvegicus JN167537/Germany/2009/Rattus norvegicus	
P(122)L	LC549185/China/2012/Rattus losea LC549184/China/2012/Rattus norvegicus JN167538/Germany/2009/Rattus norvegicus	
T(123)I	JN167538/Germany/2009/Rattus norvegicus JN167537/Germany/2009/Rattus norvegicus	
L(130)S	LC549185/China/2012/Rattus losea LC549184/China/2012/Rattus norvegicus	
Q(131)R	GU345043/Germany/2009/Rattus norvegicus LC549185/China/2012/Rattus losea LC549184/China/2012/Rattus norvegicus	
L(134)R	JN167537/Germany/2009/Rattus norvegicus LC549185/China/2012/Rattus losea LC549184/China/2012/Rattus norvegicus	
R(142)Q	LC549185/China/2012/Rattus losea	
R(142)H	JN167537/Germany/2009/Rattus norvegicus	
V(143)M	LC549185/China/2012/Rattus losea LC549184/China/2012/Rattus norvegicus	
V(143)A	JN167537/Germany/2009/Rattus norvegicus	
G(144)D	LC549185/China/2012/Rattus losea LC549184/China/2012/Rattus norvegicus JN167537/Germany/2009/Rattus norvegicus	
S(145)L	JN167537/Germany/2009/Rattus norvegicus	
N(146)S	LC549184/China/2012/Rattus norvegicus JN167538/Germany/2009/Rattus norvegicus JN167537/Germany/2009/Rattus norvegicus GU345043/Germany/2009/Rattus norvegicus	
A(147)V	LC549184/China/2012/Rattus norvegicus	
G(149)A	LC549185/China/2012/Rattus losea LC549184/China/2012/Rattus norvegicus	
P(132)L	JN167537/Germany/2009/Rattus norvegicus	
R(153)T	LC549185/China/2012/Rattus losea	
R(153)M	LC549184/China/2012/Rattus norvegicus	

S(154)W	LC549185/China/2012/Rattus losea	A(75)V	AB890374/Japan/2013/Mustela putorius furo
A(155)E	LC549184/China/2012/Rattus norvegicus		AB890001/Japan/2013/Mustela putorius furo
A(155)V	LC549185/China/2012/Rattus losea	S(97)F	LC057247/Japan/2013/Mustela putorius furo
S(156)L	LC549184/China/2012/Rattus norvegicus	T(99)I	LC177789/USA/2016/Mustela putorius furo
C(158)S	JN167537/Germany/2009/Rattus norvegicus		AB890374/Japan/2013/Mustela putorius furo
C(158)F	JN167538/Germany/2009/Rattus norvegicus		AB890001/Japan/2013/Mustela putorius furo
I(159)T	JN167537/Germany/2009/Rattus norvegicus		LC177792/USA/2016/Mustela putorius furo
	LC549185/China/2012/Rattus losea		LC177791/USA/2016/Mustela putorius furo
	JN167538/Germany/2009/Rattus norvegicus		LC177790/USA/2016/Mustela putorius furo
	JN167537/Germany/2009/Rattus norvegicus		LC177789/USA/2016/Mustela putorius furo
C(161)S	LC549185/China/2012/Rattus losea		LC177788/USA/2016/Mustela putorius furo
	LC549184/China/2012/Rattus norvegicus		LC057247/Japan/2013/Mustela putorius furo
T(162)M	GU345043/Germany/2009/Rattus norvegicus	R(104)H	AB890374/Japan/2013/Mustela putorius furo
R(168)K	JN167538/Germany/2009/Rattus norvegicus		AB890001/Japan/2013/Mustela putorius furo
	JN167537/Germany/2009/Rattus norvegicus		LC057247/Japan/2013/Mustela putorius furo
	LC549184/China/2012/Rattus norvegicus		LC177792/USA/2016/Mustela putorius furo
Q(172)L	LC549185/China/2012/Rattus losea	T(106)M	LC177791/USA/2016/Mustela putorius furo
	LC549184/China/2012/Rattus norvegicus		LC177790/USA/2016/Mustela putorius furo
Q(172)R	JN167537/Germany/2009/Rattus norvegicus		LC177789/USA/2016/Mustela putorius furo
V(177)A	LC549185/China/2012/Rattus losea		LC177788/USA/2016/Mustela putorius furo
	LC549184/China/2012/Rattus norvegicus		LC177792/USA/2016/Mustela putorius furo
L(182)M	JN167538/Germany/2009/Rattus norvegicus	G(119)E	LC177791/USA/2016/Mustela putorius furo
	LC549185/China/2012/Rattus losea		LC177790/USA/2016/Mustela putorius furo
	LC549184/China/2012/Rattus norvegicus		LC177789/USA/2016/Mustela putorius furo
L(182)T	JN167537/Germany/2009/Rattus norvegicus		LC177788/USA/2016/Mustela putorius furo
L(182)S	GU345043/Germany/2009/Rattus norvegicus	Q(122)L	AB890374/Japan/2013/Mustela putorius furo
			AB890001/Japan/2013/Mustela putorius furo
Dataset III			LC057247/Japan/2013/Mustela putorius furo
		I(124)T	AB890374/Japan/2013/Mustela putorius furo
			AB890001/Japan/2013/Mustela putorius furo
T(8)I	LC057247/Japan/2013/Mustela putorius furo		LC057247/Japan/2013/Mustela putorius furo
V(11)A	AB890374/Japan/2013/Mustela putorius furo	R(134)Q	LC177792/USA/2016/Mustela putorius furo
	LC177792/USA/2016/Mustela putorius furo		LC177791/USA/2016/Mustela putorius furo
	LC177791/USA/2016/Mustela putorius furo		LC177790/USA/2016/Mustela putorius furo
	LC177790/USA/2016/Mustela putorius furo		LC177789/USA/2016/Mustela putorius furo
	LC177789/USA/2016/Mustela putorius furo		LC177788/USA/2016/Mustela putorius furo
	LC177788/USA/2016/Mustela putorius furo	R(153)Q	AB890374/Japan/2013/Mustela putorius furo
	LC057247/Japan/2013/Mustela putorius furo		AB890001/Japan/2013/Mustela putorius furo
K(14)R	AB890374/Japan/2013/Mustela putorius furo		LC177792/USA/2016/Mustela putorius furo
	AB890001/Japan/2013/Mustela putorius furo		LC177791/USA/2016/Mustela putorius furo
	LC177792/USA/2016/Mustela putorius furo		LC177790/USA/2016/Mustela putorius furo
	LC177791/USA/2016/Mustela putorius furo		LC177789/USA/2016/Mustela putorius furo
	LC177790/USA/2016/Mustela putorius furo		LC177788/USA/2016/Mustela putorius furo
	LC177789/USA/2016/Mustela putorius furo	F(161)S	LC057247/Japan/2013/Mustela putorius furo
	LC177788/USA/2016/Mustela putorius furo		LC177792/USA/2016/Mustela putorius furo
	LC057247/Japan/2013/Mustela putorius furo		LC177791/USA/2016/Mustela putorius furo
R(16)Q	AB890374/Japan/2013/Mustela putorius furo		LC177790/USA/2016/Mustela putorius furo
	AB890001/Japan/2013/Mustela putorius furo		LC177789/USA/2016/Mustela putorius furo
	LC177792/USA/2016/Mustela putorius furo		LC177788/USA/2016/Mustela putorius furo
	LC177791/USA/2016/Mustela putorius furo	S(169)L	LC177792/USA/2016/Mustela putorius furo
	LC177790/USA/2016/Mustela putorius furo		LC177791/USA/2016/Mustela putorius furo
	LC177789/USA/2016/Mustela putorius furo		LC177790/USA/2016/Mustela putorius furo
	LC177788/USA/2016/Mustela putorius furo		LC177789/USA/2016/Mustela putorius furo
	LC057247/Japan/2013/Mustela putorius furo		LC177788/USA/2016/Mustela putorius furo
Q(18)R	AB890374/Japan/2013/Mustela putorius furo	T(174)I	AB890374/Japan/2013/Mustela putorius furo
	AB890001/Japan/2013/Mustela putorius furo		AB890001/Japan/2013/Mustela putorius furo
	LC177792/USA/2016/Mustela putorius furo		LC177792/USA/2016/Mustela putorius furo
	LC177791/USA/2016/Mustela putorius furo		LC177791/USA/2016/Mustela putorius furo
	LC177790/USA/2016/Mustela putorius furo		LC177790/USA/2016/Mustela putorius furo
	LC177789/USA/2016/Mustela putorius furo		LC177789/USA/2016/Mustela putorius furo
	LC177788/USA/2016/Mustela putorius furo		LC177788/USA/2016/Mustela putorius furo
	LC057247/Japan/2013/Mustela putorius furo		LC057247/Japan/2013/Mustela putorius furo
H(31)P	AB890374/Japan/2013/Mustela putorius furo	A(175)V	JN998606/Netherlands/2010/Mustela putorius
	AB890001/Japan/2013/Mustela putorius furo	M(182)T	AB890374/Japan/2013/Mustela putorius furo
	LC177792/USA/2016/Mustela putorius furo		AB890001/Japan/2013/Mustela putorius furo
	LC177791/USA/2016/Mustela putorius furo		LC057247/Japan/2013/Mustela putorius furo
	LC177790/USA/2016/Mustela putorius furo		
	LC177789/USA/2016/Mustela putorius furo		
	LC177788/USA/2016/Mustela putorius furo		
	LC057247/Japan/2013/Mustela putorius furo		
N(49)S	AB890374/Japan/2013/Mustela putorius furo		
	AB890001/Japan/2013/Mustela putorius furo		
	LC177792/USA/2016/Mustela putorius furo		
	LC177791/USA/2016/Mustela putorius furo		
	LC177790/USA/2016/Mustela putorius furo		
	LC177789/USA/2016/Mustela putorius furo		
	LC177788/USA/2016/Mustela putorius furo		
	LC057247/Japan/2013/Mustela putorius furo		
N(53)S	AB890374/Japan/2013/Mustela putorius furo		
	AB890001/Japan/2013/Mustela putorius furo		
	LC177792/USA/2016/Mustela putorius furo		
	LC177791/USA/2016/Mustela putorius furo		
	LC177790/USA/2016/Mustela putorius furo		
	LC177789/USA/2016/Mustela putorius furo		
	LC177788/USA/2016/Mustela putorius furo		
	LC057247/Japan/2013/Mustela putorius furo		

*The number in the parentheses indicated the location of amino acid in its protein.

Table 3: Codon values of different parameters for ORF4 genomes in different hosts

Parameters	Transitional	Substituted
ma	03	3 C
t	68	5 C
ret	22	0 C

S: Number of segregating sites
 π: Nucleotide diversity
 R: Transition/Transversion bias

Table S1: Codon positions with entropy score in the ORF4 genomes of Hepatitis E viruses

No	Codon positions with entropy score
Human	

1	Position 66: 0.50040
Rat	
1	Position 2: 0.41012
2	Position 11: 0.68291
3	Position 18: 0.68291
4	Position 29: 0.59827
5	Position 31: 0.41012
6	Position 32: 0.41012
7	Position 35: 0.68291
8	Position 38: 0.59827
9	Position 39: 0.59827
10	Position 40: 1.07899
11	Position 44: 0.68291
12	Position 45: 0.41012
13	Position 53: 0.68291
14	Position 54: 0.41012
15	Position 64: 0.68291
16	Position 71: 0.68291
17	Position 73: 0.68291
18	Position 75: 0.59827
19	Position 76: 0.68291
20	Position 80: 0.59827
21	Position 84: 0.68291
22	Position 89: 0.41012
23	Position 94: 0.41012
24	Position 98: 0.59827
25	Position 99: 0.59827
26	Position 102: 0.41012
27	Position 103: 0.68291
28	Position 104: 0.68291
29	Position 105: 0.41012
30	Position 106: 0.41012
31	Position 112: 0.68291
32	Position 115: 0.41012
33	Position 116: 0.59827
34	Position 118: 0.41012
35	Position 119: 0.95570
36	Position 120: 0.59827
37	Position 121: 0.68291
38	Position 122: 0.68291
39	Position 123: 0.59827
40	Position 130: 0.68291
41	Position 131: 0.68291
42	Position 134: 0.68291
43	Position 142: 0.79631
44	Position 143: 0.95570
45	Position 144: 0.68291
46	Position 145: 0.41012
47	Position 146: 0.68291
48	Position 147: 0.41012
49	Position 148: 0.00000
50	Position 149: 0.59827
51	Position 152: 0.41012
52	Position 153: 0.79631
53	Position 154: 0.59827
54	Position 155: 0.79631
55	Position 156: 0.41012
56	Position 158: 0.79631
57	Position 159: 0.68291
58	Position 161: 0.59827
59	Position 162: 0.41012
60	Position 168: 0.68291
61	Position 172: 0.95570
62	Position 177: 0.59827
63	Position 182: 1.27703
Ferret	
1	Position 8: 0.32508
2	Position 11: 0.50040
3	Position 14: 0.50040
4	Position 16: 0.50040
5	Position 18: 0.50040
6	Position 31: 0.50040
7	Position 49: 0.50040
8	Position 53: 0.50040
9	Position 75: 0.61086
10	Position 97: 0.32508
11	Position 99: 0.50040
12	Position 104: 0.61086
13	Position 106: 0.69315
14	Position 119: 0.69315

15	Position 122: 0.61086
16	Position 124: 0.61086
17	Position 134: 0.69315
18	Position 153: 0.50040
19	Position 161: 0.69315
20	Position 169: 0.69315
21	Position 174: 0.50040
22	Position 175: 0.32508
23	Position 182: 0.61086

Table S2: Secondary structure elements prediction by SOPMA

S. No.	Secondary structure elements	Values (%)			
		LC057248	KU168733	JN167538	LC177791
1	Alpha helix	32.24	8.23	39.89	37.70
2	3 ₁₀ -helix	0.00	0.00	0.00	0.00
3	Pi helix	0.00	0.00	0.00	0.00
4	Beta bridge	0.00	0.00	0.00	0.00
5	Extended strand	26.23	21.52	24.04	20.77
6	Beta turn	4.92	8.86	5.46	4.37
7	Bend region	0.00	0.00	0.00	0.00
8	Random coil	36.61	61.39	30.60	37.16
9	Ambiguous states	0.00	0.00	0.00	0.00
10	Other states	0.00	0.00	0.00	0.00

Results and Discussion

To study the structure and function of protein, *in-silico* analyses have become a very valuable method [27]. Recently, analysis on proteins using *in-silico* tools has provided a huge contribution to the field of computational biology in elucidating the protein’s functional and structural aspects [28, 29]. In this context, we exploited different computational tools to reveal significant information on the ORF4 proteins of HEV.

Analysis of mutations in ORF4 protein genes

RNA viruses mutate at a very high rate, i.e., 10⁶ to 10⁴ new base substitutions per nucleotide per cell. Additionally, it has been well documented that virus with single-stranded genome appears to mutate faster than double-stranded viruses [30]. For mutational analysis, the sequence NC_038504/ Germany/2009, KU168733/India/2013 and JN998607/Netherland/2010 was used as a reference genome for dataset I, II and III respectively. The predicted mutations in the ORF4 genomes for datasets, i.e., Human, Rat and Ferret are summarized in **Table 2**. Our mutational analysis mostly showed changes in ORF4 genes, which corresponded to both synonymous and non-synonymous mutations (**Figure 1**). Thus, it can be interpreted that HEV also exhibits a high degree of genetic variation like other RNA viruses, due to viral RNA polymerase non-proofreading activity, rapid rates of replication, immense population size, and immunological pressure [30]. Thus, the previous hypothesis suggesting high mutation rates in RNA viruses substantiate our findings.

Analysis of entropy in ORF4 protein sequences

The entropy is one useful method of quantification of diversity in amino acid sequences [31]. Structurally or functionally important amino acid variations are correlated with high scoring entropy values [32]. The entropy analysis revealed a total of 1, 63 and 23 sites were identified in datasets I, II and III for Human, Rat and Ferret respectively (**S1 Table**). The entropy percentages for ORF4 genomes are as follow: Human: 0.006% (1/159), Rat: 0.342% (63/184) and Ferret: 0.125% (23/184) respectively. Therefore, ORF4 genomes in rat observed the largest variation followed by ferret genomes and human genomes had the least variation (**Figure 2**). However, further thorough experimental investigations in conjunction with other studies (site directed mutagenesis) are mandatory to establish relationships between the reported mutations and their corresponding functional changes. Moreover, detailed insights into the mechanism of these strains are needed to confirm their pathogenicity and zoonotic potential.

Analysis of positive selection in ORF4 protein genes

Gene selective pressure for genes was estimated using the Tajima's Neutrality Test. The results suggested that genes comprising dataset I was found to be under purifying selection as indicated by negative *D* value, i.e., -1.093. However, dataset II and III consisted of genes under positive selection as indicated by positive *D* values, i.e., 0.861 and 0.554 for Rat and Ferret respectively. The selection pressure revealed the prevalence of positively selected sites in datasets II and III. While prevalence of purifying selection in Human dataset was observed. This suggests that the ORF4 region evolution is mainly driven by positive selection in Rat and Ferret. Thus, prevalence of non-synonymous mutations with high entropy scores corresponding to positively

selected sites in Rat and Ferret datasets suggested high variability in these ORF4 protein genes.

Estimation of codon degeneracy patterns between hosts in ORF4 protein genes

The variation in codon properties was examined due to the codon degeneracy that was maintained in ORF4 protein genes.

Nucleotide diversity (π): The least nucleotide diversity in codon pattern was observed in Humans, and maximum divergence was found in Rat. The value of Ferret was intermediate between Human and Rat. The codon patterns followed the order of nucleotide divergence in the order Rat > Ferret > Human (Table 3).

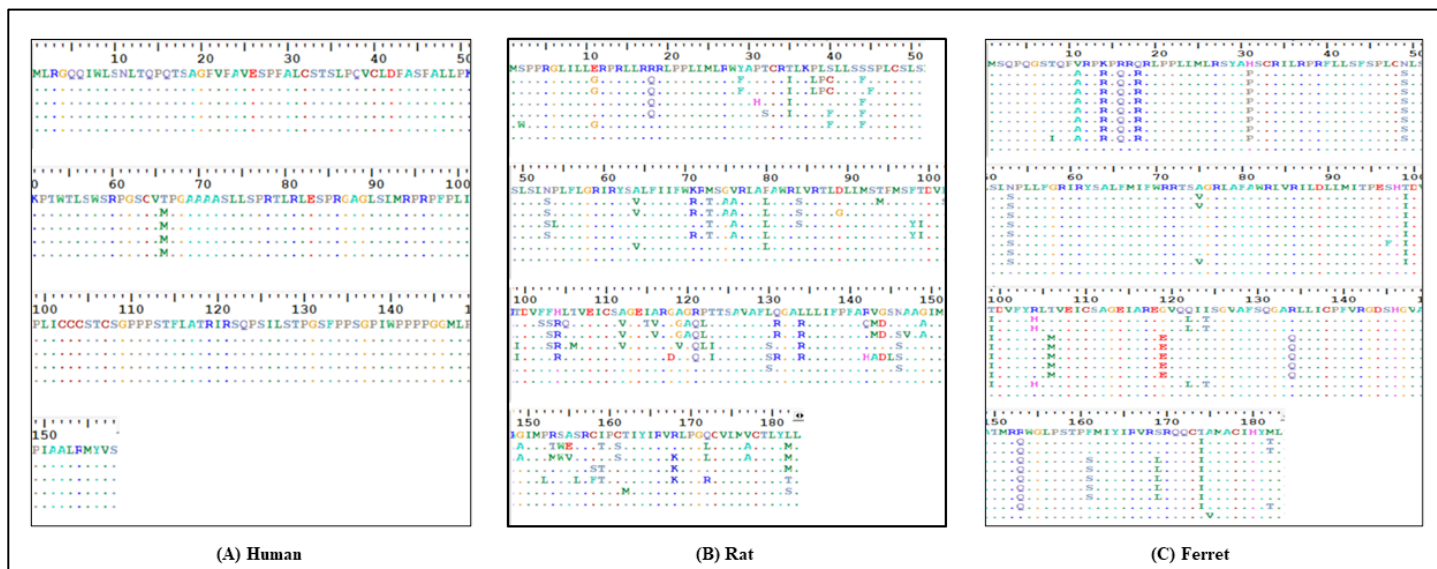


Figure 1 Alignment showing the comparative analysis of amino acid sequences in the host organisms (A) Human; (B) Rat and (C) Ferret. The substitutions are shown by amino acid symbols at the respective positions and similarities are represented by the dots.

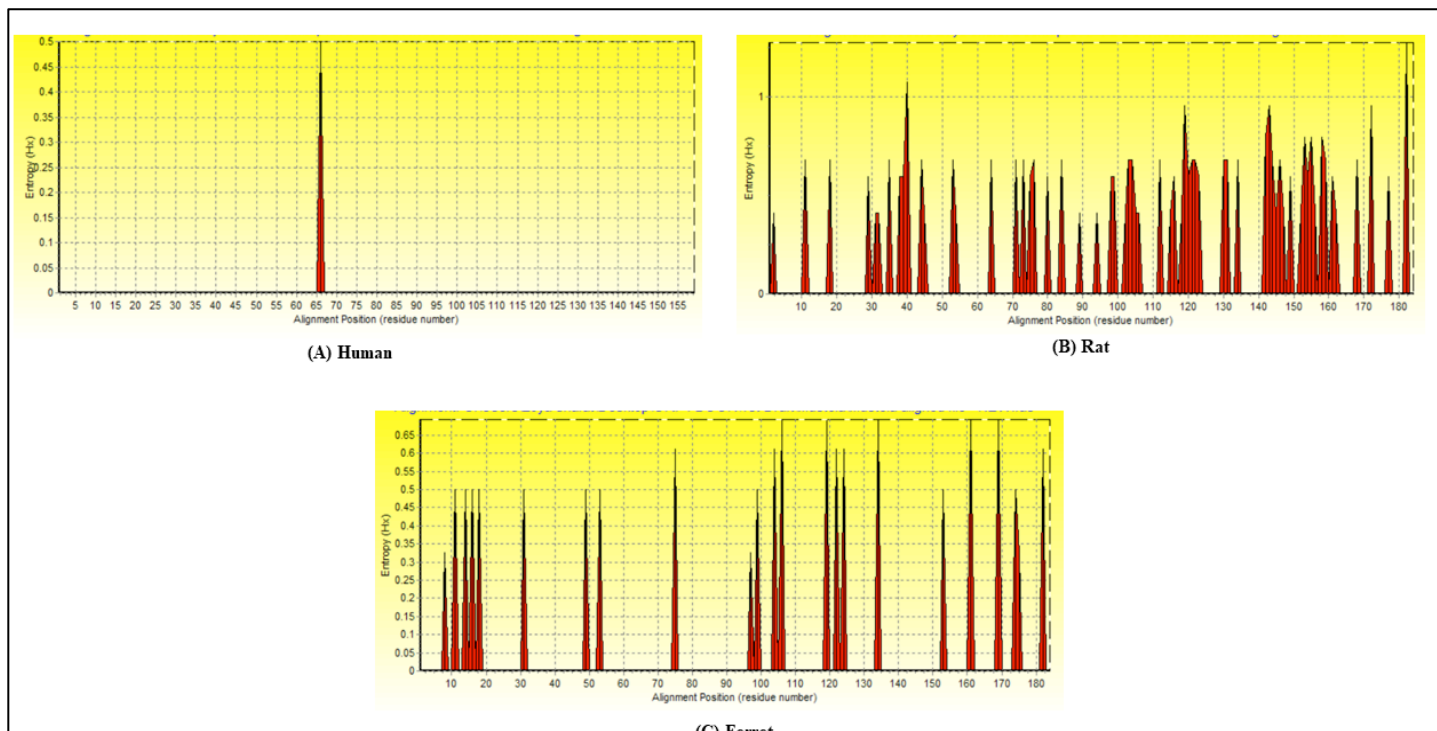


Figure 2 Comparative analysis of entropy of amino acid sequences in ORF4 in the host organisms (A) Human; (B) Rat and (C) Ferret.

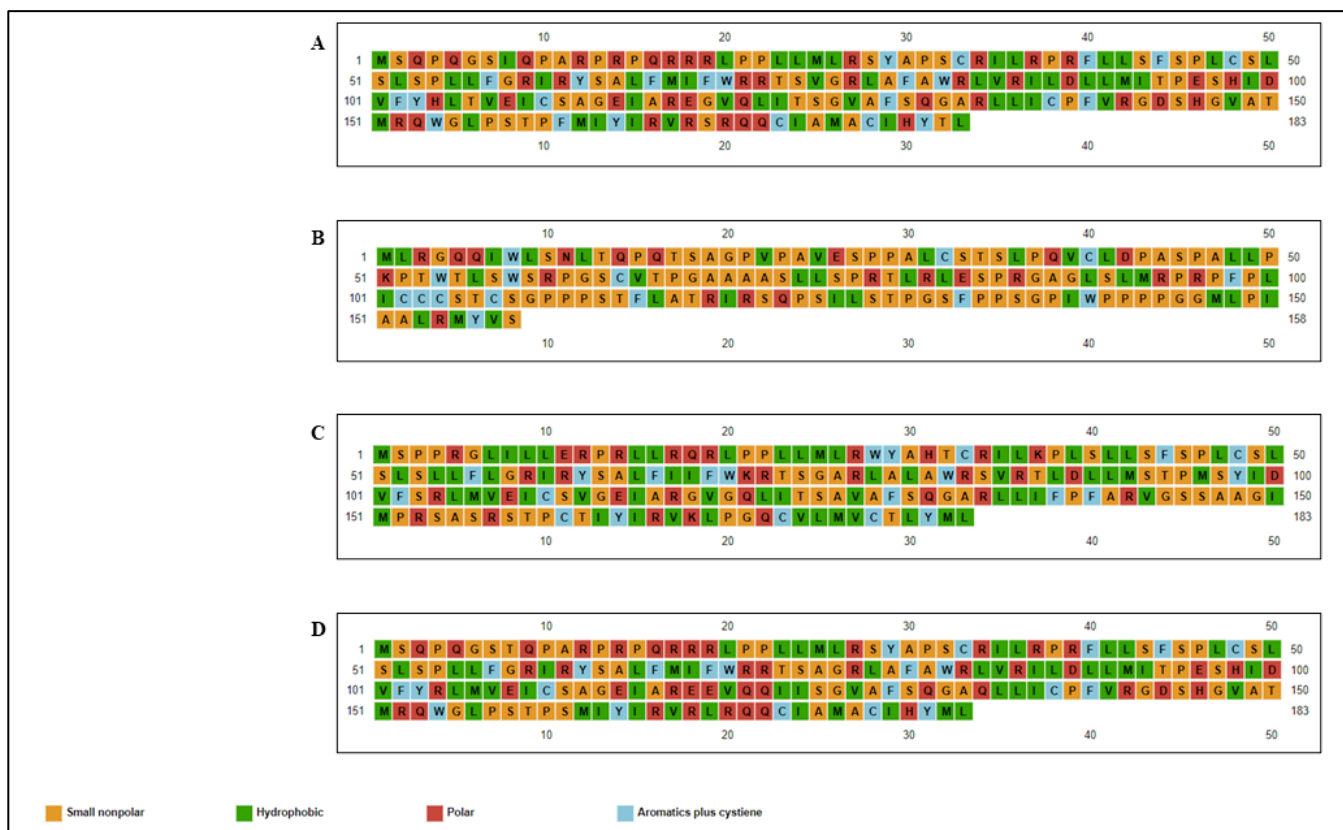


Figure 3 Analysis of amino acid composition in ORF4. Representation of small non-polar, hydrophobic, polar, aromatic plus cysteine amino acid residues in different ORF4 sequences. (A) LC057248 (HEV); (B) KU168733 (Human); (C) JN167538 (Rat); and (D) LC177791 (Ferret). The analysis was conducted using the PSIPRED.

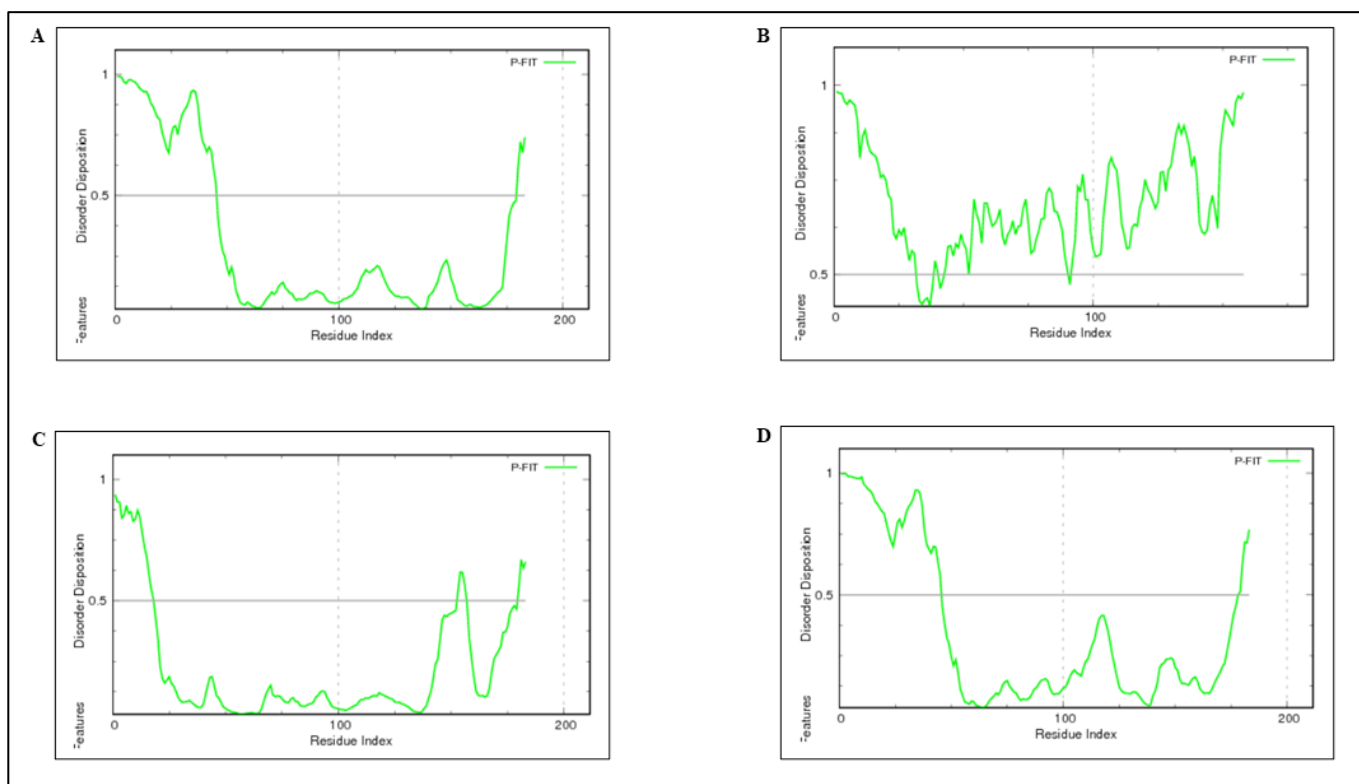


Figure 4 Analysis of intrinsic disorder predisposition of ORF4. Representation of intrinsic disorder profile of ORF4 in different sequences (A) LC057248 (HEV); (B) KU168733 (Human); (C) JN167538 (Rat); and (D) LC177791 (Ferret). Disorder probability was calculated using DisProt (version of PONDR-FIT), where graphs represent the intrinsic disorder profiles. A threshold value of 0.5 was set to distinguish between ordered and disordered region along the genome (dashed line). Regions above the threshold are predicted to be disordered.

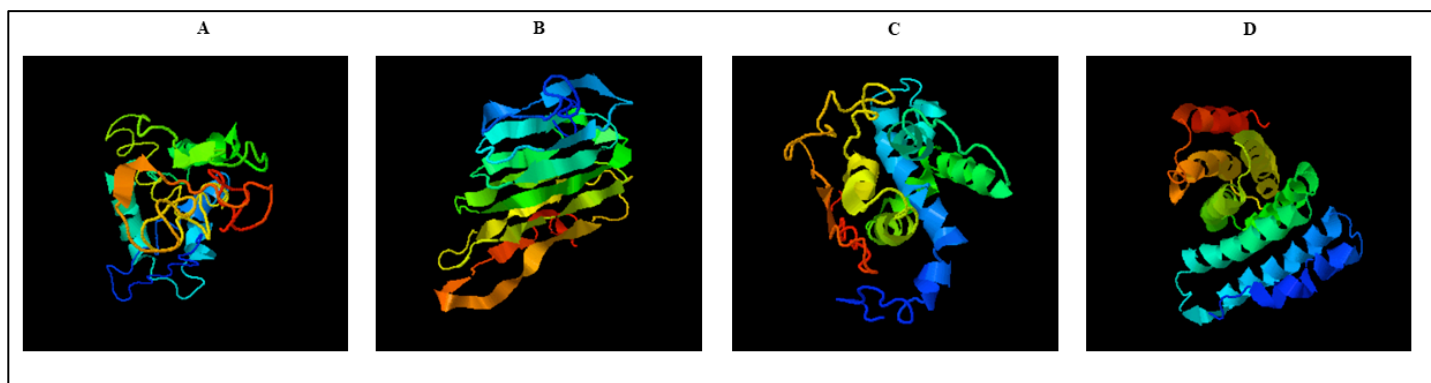


Figure 5: ORF4 with the predicted tertiary structure: (A) LC057248 (HEV); (B) KU168733 (Human); (C) JN167538 (Rat); and (D) LC177791 (Ferret). The analysis was conducted using I-TASSER web server.

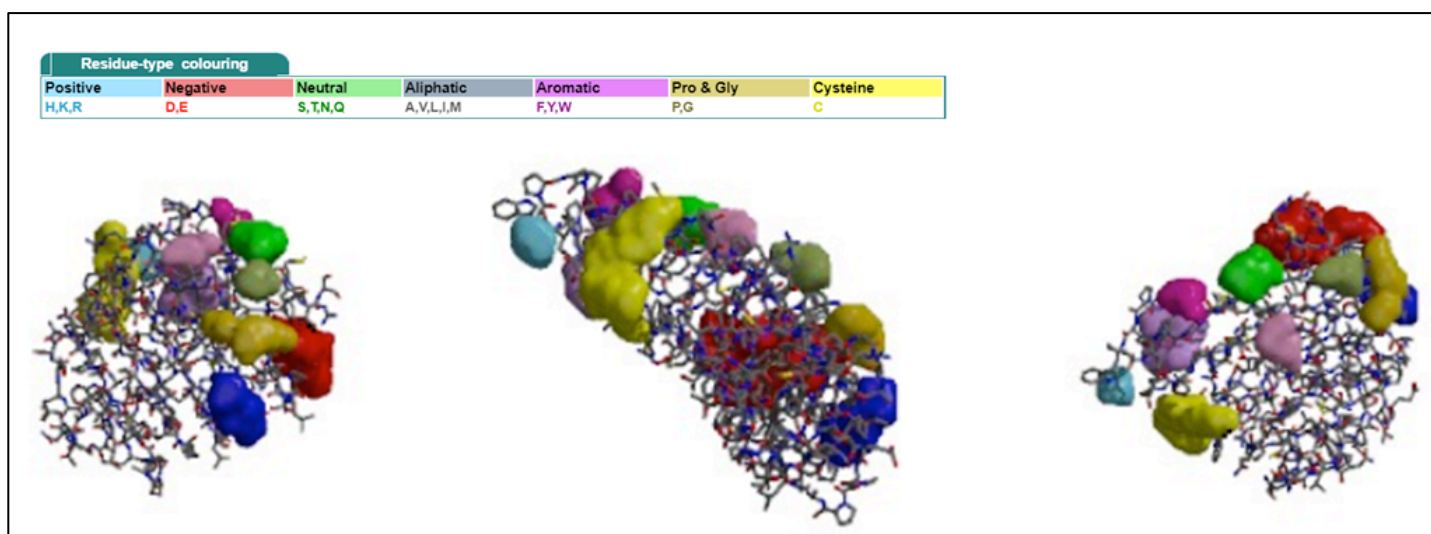


Figure S1: Representation of clefts in the predicted ORF4 3D structural mode KU168733 (Human). The analysis was conducted using PDBsum.

Number of segregating sites (S):

The estimated segregated site in the ORF4 was in accordance with the nucleotide diversity (π). The highest S was correlated with highest π value. The codon patterns followed the order of segregation sites in the order Rat > Ferret > Human (Table 3).

Transitions more common than transversions (R):

The estimated transition/transversion bias for the hosts ranges from 0.3 to 5.5 in the ORF4 region. Rate of occurrence of transitional substitutions were much greater than the rate of transversion substitutions in all the natural hosts (Table 3). Higher transition/transversion ratio values in ORF4 region also reveals that less diversity in the amino acid composition due to less transversions, as more transversions which result in substantial dissimilar chemical composition [33]. Our results are in accordance with the previous study on HEV that suggested high transition to transversion ratio [34]. The phenomenon is mainly attributed to two mutually non-exclusive hypotheses: the mutational hypothesis and the selective hypothesis. The mutational hypothesis posits that transition rates are higher than the transversion mutation rates in both the coding and non-coding sequences [35, 36]. The selective hypothesis holds that natural selection disfavors transversions [30, 37]. Thus, our investigation showing biasness towards transition suggests that transitional mutations are more favored than transversions in the ORF4 region, which supports earlier mentioned hypotheses [34]. Thus, it can be interpreted that both mutation and natural selection influenced the ORF4 genomes. This is consistent with

earlier report that revealed the co-existence of mutation-selection balance in RNA viruses [33].

Analysis of structure of ORF4 proteins

Earlier studies have revealed that a certain type of protein has been recognized with a lack of a well defined structure under physiological conditions but perform crucial biological functions. This class of proteins or protein regions are defined as intrinsically disordered proteins (IDPs) or intrinsically disordered protein regions (IDRs) [13, 14]. An IDP possesses a unique feature, which enables it to interact with one to many and many to one signaling [38]. The significance of IDPs in biological functions, such as recognition, regulation, signaling, and protein-protein interaction (PPI) network control has been well documented [15]. IDPs are closely linked with human diseases (tumor, cardiovascular disease, neurodegenerative diseases, and diabetes) [16 - 20]. Due to IDPs involvement in diverse signaling and regulatory processes, strategies in drug discovery aiming at IDPs have gained momentum [15, 22]. Therefore, these IDPs due to their unique structures act as potential targets in drug designing. Furthermore, IDPs are usually hub proteins in PPI networks, and PPIs are potential sources for drug targets. Thus, in this study we have examined the sequence and structure of ORF4 protein in order to reveal their prime features as a potential for drug target molecule. IDPs can be easily predicted by bioinformatical methods due to their peculiar amino acid composition [39 - 43]. Dunker and colleagues [9] categorized amino acids into three groups based on their composition enrichment in ordered and disordered segments, i.e., the order-promoting group (C, W, Y, I, F, V and L), the disorder-promoting group (M, K, R, S, Q, P and E)

and the neutral group (A, G, H, T, N and D) [44]. Initially, we performed a sequence-based comprehensive analysis of ORF4 proteins (LC057248, KU168733, JN167538 and LC177791) in terms of amino acid composition to elucidate their functional properties (**Figure 3**). Our results clearly revealed that all the ORF4 sequences were enriched in characteristic disorder-promoting residues (Arg, Pro and Ser) and neutral residues (Ala, Gly and Thr). Additionally, abundance of high proportion of structure-breaking residues (Gly and Pro) has been suggested that the protein is an IDP [45]. Also, the largest fractional change between the ordered and disordered protein is exhibited by Pro [46]. Thus, abundance of Pro amino acid residue in the ORF4 protein (KU168733), clearly indicated that the ORF4 protein (KU168733) particularly contains significant fraction of intrinsic disorder in comparison to other ORF4 proteins LC057248, JN167538 and LC177791. After the initial primary structure analysis, the secondary structure elements were determined that showed the presence of all three major contents including alpha helix, beta-strand and coils (**S2 Table**). However, it was evident from our results that the ORF4 protein obtained from human (KU168733) was characterized with prevalence of coils and lack of secondary structure elements (helix and sheet) in comparison to other ORF4 proteins (**S2 Table**). Protein-protein interactions (PPIs) are considered as potential sources for drug targets [22]. Intrinsic disorder is utilized in protein-protein interactions: namely, one disordered region binding to many partners and many disordered regions binding to one partner [38]. Therefore, we analyzed the predisposition of intrinsic disorder of ORF4 proteins. Based on predicted percentage of intrinsic disorder (PPID) in ORF4, the ORF4 sequences were classified into different protein variants: Ordered proteins (ORDPs); Intrinsically disordered protein regions (IDPRs); and Intrinsic disordered proteins (IDPs) [47]. The first category ORDP includes protein sequences, which have PPID less than 10%. The IDPR category includes protein sequences having PPID 10 - 30%. Lastly, the IDP category includes protein sequences, which are predicted to have PPID more than 30%. Thus, based on PPID, the ORF4 protein sequences considered in the study were categorized into different variants. The ORF4 proteins LC057248 (HEV) and LC177791 (ferret) were categorized into the intrinsically disordered protein regions (IDPRs), as they consisted of PPID in the range between 10% to 30% (**Figure 4A and 4D**). The ORF4 protein JN167538 (rat) was categorized into ORDPs, as it consisted of less than 10% of PPID (**Figure 4C**). The ORF4 protein KU168733 (human) was categorized into the IDPs as it consisted more than 30% of PPID. It was observed that the major portion of the polypeptide chain of the ORF4 protein KU168733 was highly disordered, revealing it as an IDP (**Figure 4B**). Thus, taken altogether, beginning from the initial sequence analysis, secondary structure element up to fraction of intrinsic disorder content, it is clearly revealed that the ORF4 protein obtained from host human possesses the attributes of an IDP. IDPs perform significant roles in recognition, regulation, signaling, and protein-protein interaction (PPI) network control, thus are considered as potential targets in structure-based drug designing. Moreover, IDPs generally represent themselves as hub proteins in PP1 networks, and PPIs are potential sources for drug targets [15, 22].

Furthermore, the generated 3D ORF4 protein models were comparatively visualized (**Figure 5**). Compared with other ORF4 proteins, i.e., JN167538 (ORDP), LC057248 and LC177791 (IDPRs), the ORF4 protein KU168733 (IDP) 3D model possessed a highly flexible and random coiled-like structure (**Figure 5B**), which shows consistency with the previous report suggesting IDPs fail to arrange into a definite 3D structure under physiological conditions due to increased level of disorder-promoting residues [44]. Thus, out of several models, the obtained model

from host Human can be considered as a reliable drug target due to its characteristic highly disordered (IDP) structure [15 - 20, 22]. Additionally, identification of clefts, tunnels and pores accessible to ligand molecules is essential in the context of structure-based drug design process [48, 49]. Thus, the modelled structure of ORF4 protein (KU168733) was scrutinized using PDBsum analysis to reveal the presence of binding sites. Interestingly, the modelled ORF4 protein revealed the presence of 10 clefts (**S1 Figure**), which determines their interaction with other molecules [50]. Clefts or pockets present on protein's surface are sizeable depressions that have tendency to be enzyme active sites [48]. Thus, to sum up our observations it can be interpreted that ORF4 protein (KU168733), due to its characteristics of an IDP, i.e., prevalence of Gly, Pro and Ser, lack of secondary structure with the predominance of coils, in addition to presence of several clefts, suggest its commitment towards interaction with other target molecules. Thus, it can be considered as a reliable drug target.

Conclusion:

This novel study was aimed to collect information and discusses in the ORF4 of HEV. It provided detailed analysis on the occurrence of genomic diversity in the ORF4 protein genes of HEV. Further, the ORF4 protein of HEV was analyzed at different structural levels to shed light on its putative functions. Our presented results on function and structure of HEV ORF4 are theoretical hypotheses. Therefore, validations involving ORF4 structure by both computational and experimental approaches are further required.

Funding: Not applicable

Authors' contributions:

SP conceptualized the research. SP and ZS designed the manuscript. ZS was a major contributor in writing the manuscript and performed the biocomputational analysis of the protein. KP and AA proofread the manuscript. All the authors read and approved the final manuscript.

Acknowledgements:

The authors would like to acknowledge Maulana Azad National Fellowship (MANF), University Grant Commission (UGC), Council of Scientific and Industrial Research (CSIR) (37(1697)17/EMR-II) and Central Council for Research in Unani Medicine (CCRUM), Ministry of Ayurveda, Yoga and Neuropathy, Unani, Siddha and Homeopathy (AYUSH) (F.No.3-63/2019-CCRUM/Tech) supported by the Government of India.

References:

- [1] Kumar S et al. *Int J Infect Dis*. 2013 **17**:e228 [PMID: 23313154].
- [2] Khuroo MS c Khuroo MS, 2016 *J Viral Hepat*. **23**:68 [].
- [3] Takahashi M et al. *J Clin Microbiol*. 2010 **48**:1112 [PMID: 20107086].
- [4] Tam AW et al. *Virology* 1991 **185**:120 [PMID: 1926770].
- [5] Kenney SP & Meng X], *Cold Spring Harb Perspect Med*. 2019 **9**:a031724 [PMID: 29530948].
- [6] Nair VP et al. *PLoS Pathog*. 2016 **12**:e1005521 [PMID: 27035822].
- [7] Subramani C et al. *MSystems* 2018 **3**:e00135 [PMID: 29404423].
- [8] Li TC et al. *Emerg Infect Dis*. 2014 **20**:709 [PMID: 24655541].
- [9] Kobayashi T et al. *Virus Res*. 2018 **249**:16 [PMID: 29471051].
- [10] Grad YH & Lipsitch M, *Genome Biol*. 2014 **15**:1 [PMID: 25418119].
- [11] Haffar S et al. *Liver Int*. 2015 **35**:311.

- [12] Wang J et al. *Int J Mol Sci.* 2011 **12**:3205 [PMID: 21686180].
- [13] Schweers O et al. *J Biol Chem.* 1994 **269**: 24290 [PMID: 7929085].
- [14] Uversky VN, *Protein Sci.* 2002 **11**:739 [PMID: 11910019].
- [15] Dafforn TR & Smith CJ, *EMBO Rep.* 2004 **5**:1046 [PMID: 15520805].
- [16] Goh KI et al. *Proc Natl Acad Sci USA* 2007 **104**:8685 [PMID: 17502601].
- [17] Midic, U et al. *BMC Genomics* 2009, **10**:S12 [PMID: 19594871].
- [18] Midic U et al. *Protein Pept Lett.* 2009 **16**:1533 [PMID: 20001916].
- [19] Raychaudhuri S et al. *PLoS One* 2009, **4**:e5566 [PMID: 19440375].
- [20] Shimizu K & Toh H, *J Mol Biol.* 2009 **392**:1253 [PMID: 19660471].
- [21] Dunker AK et al. *Curr Opin Struct Biol.* 2008 **18**:756 [PMID: 18952168].
- [22] Metallo SJ, *Curr Opin Chem Biol.* 2010 **14**:481 [PMID: 20598937].
- [23] Hall TA. BioEdit v 7.2. 3. Biological sequence alignment editor for Win 95/98/NT/2K/XP7.
- [24] Tamura K et al. *Mol Evol Genet.* 2013 **30**:2725 [PMID: 24132122].
- [25] Zhang Y, *BMC Bioinformatics* 2008 **9**:1 [PMID: 18215316].
- [26] Laskowski RA, *J Biomol NMR* 1996 **8**:477 [PMID: 9008363].
- [27] Santhoshkumar R & Yusuf A, *J Genet Eng Biotechnol.* 2020 **18**:1 [PMID: 32617758].
- [28] Verma A et al. *Comput Biol Chem.* 2016 **60**:53 [PMID: 26672917].
- [29] Dutta B et al. *J Genet Eng Biotechnol.* 2018 **16**:749 [PMID: 30733796].
- [30] Sanjuán R & Domingo-Calap P, *Reference Module in Life Sciences* 2019 [PMCID: PMC7157443].
- [31] Pan K & Deem MW, *J R Soc Interface* 2011 **8**:1644 [PMID: 21543352].
- [32] Wang K & Samudrala R *BMC Bioinformatics* 2006 **7**:1 [PMID: 16916457].
- [33] Zhang J, *J Mol Evol.* **50**:56 2000 [PMID: 10654260].
- [34] Lyons DM & Lauring AS, *Mol Biol Evol.* 2017 **34**:3205 [PMID: 29029187].
- [35] Jiang C & Zhao Z, *Genomics* 2006 **88**:527 [PMID: 16860534].
- [36] Zhang Z & Gerstein M, *Nucleic Acids Res.* 2003 **31**:5338 [PMID: 12954770].
- [37] Vogel F & Kopun M, *J Mol Evol.* 1977 **9**:159 [PMID: 864721].
- [38] Uversky VN, *Protein J.* 2009 **28**:305 [PMID: 19768526].
- [39] Campen A et al. *Protein Pept Lett.* 2008 **15**:956 [PMID: 18991772].
- [40] Vacic V et al. *BMC Bioinform.* 2007 **8**:211 [PMID: 17578581].
- [41] Williams RM et al. *Pac Symp Biocomput.* 2001 **2001**:89 [PMID: 11262981].
- [42] Romero P et al. *Proteins* 2001 **42**:38 [PMID: 11093259].
- [43] Garner E et al. *Genome Inform Ser Workshop Genome Inform.* 1998 **9**:201 [PMID: 11072336].
- [44] Radivojac, P et al. *Biophys J.* 2007 **92**:1439 [PMID: 17158572].
- [45] Dunker AK et al. *J Mol Graph* 2001 **19**:26 [PMID: 11381529].
- [46] Campen A et al *Protein Pept Lett.* 15:956 [PMID: 18991772].
- [47] Rajagopalan K et al. *J Cell Biochem.* 2011 **112**:3256 [PMID: 21748782].
- [48] Mbarek A et al. *Molecules* 2019 **24**:1803 [PMID: 31075983].
- [49] Marques SM et al. *Med Res Rev* 2017 **37**:1095 [PMID: 27957758].
- [50] Coleman RG & Sharp KA, *J Mmol Biol.* 2006 **362**:441 [PMID: 16934837].


Edited by P Kanguane

Citation: Shafat et al. Bioinformatics 17(9): 818-828 (2021)

License statement: This is an Open Access article which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. This is distributed under the terms of the Creative Commons Attribution License

Articles published in BIOINFORMATION are open for relevant post publication comments and criticisms, which will be published immediately linking to the original article for FREE of cost without open access charges. Comments should be concise, coherent and critical in less than 1000 words.

Agro Informatics Society



BIOMEDICAL INFORMATICS

BMIS

AGIS

OPEN ACCESS GOLD

Biomedical Informatics Society

since 2005

BIOINFORMATION
Discovery at the interface of physical and biological sciences

indexed in



PMC PubMed



INDEXED IN EMERGING SOURCES CITATION (Web of Science) WEB OF SCIENCE

EBSCO



Crossref

Web of Science Group



doi®



ResearchGate

publons

R^G