ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 18(2): 111-118 (2022)

BIOINFORMATION
Discovery at the interface of physical and biological sciences

OPEN ACCESS GOLD

BIOMEDICAL INFORMATICS

www.bioinformation.net
Volume 18(2)

Research Article

Received December 8, 2021; Revised January 3, 2022; Accepted January 3, 2022, Published February 28, 2022

DOI: 10.6026/97320630018111

Declaration on Publication Ethics:
The author's 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.

Declaration on official E-mail:
The corresponding author declares that lifetime official e-mail from their institution is not available for all authors

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

Edited by P Kangueane
Citation: Shafat *et al.* Bioinformation 18(2): 111-118 (2022)

Intrinsically disordered regions in the rodent hepevirus proteome

Zoya Shafat[1], Anwar Ahmed[2], Mohammad K. Parvez[3], Asimul Islam[1], 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; Corresponding author* - Shama Parveen

Author Contacts:
Shama Parveen - sparveen2@jmi.ac.in; Zoya Shafat - zoya179695@st.jmi.ac.in; Anwar Ahmed - anahmed@ksu.edu.sa; Asimul Islam aislam@jmi.ac.in; Mohammad K Parvez - mohkhalid@ksu.edu.sa

Abstract:
Hepatitis E virus (HEV) is the causative agent of Hepatitis E infections across the world. Intrinsically disordered protein regions (IDPRs) or intrinsically disordered proteins (IDPs) are regions or proteins that are characterized by lack of definite structure. These IDPRs or IDPs play significant roles in a wide range of biological processes, such as cell cycle regulation, control of signaling pathways, etc. IDPR/IDP in proteins is associated with the virus's pathogenicity and infectivity. The prevalence of IDPR/IDP in rat HEV proteome remains undetermined. Hence, we examined the unstructured/disordered regions of the open reading frame (ORF) encoded proteins of rat HEV by analyzing the prevalence of intrinsic disorder. The intrinsic disorder propensity analysis showed that the different ORF proteins consisted of varying fraction of intrinsic

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 18(2): 111-118 (2022)

disorder. The protein ORF3 was identified with maximum propensity for intrinsic disorder while the ORF6 protein had the least fraction of intrinsic disorder. The analysis revealed ORF6 as a structured protein (ORDP); ORF1 and ORF4 as moderately disordered proteins (IDPRs); and ORF3 and ORF5 as highly disordered proteins (IDPs). The protein ORF2 was found to be moderately as well as highly disordered using different predictors, thus, was categorized into both IDPR and IDP. Such disordered regions have important roles in pathogenesis and replication of viruses.

Keywords:
Rat hepevirus; open reading frame (ORF); intrinsic disorder; structured protein; moderately disordered protein; highly disordered protein; ordered protein; intrinsically disordered protein (IDP); intrinsically disordered protein region (IDPR).

Background:

Hepatitis E is inflammation of the liver which is caused by the Hepatitis E virus (HEV) **[1]**. Worldwide, about 20 million HEV infections and 3.3 million symptomatic hepatitis E cases occur annually which results in 44,000 deaths **[2]**. HEV is of the family *Hepeviridae* and belongs to the genus *Orthohepevirus* **[3]**. The genome of HEV is a single stranded positive-sense RNA (7.2 kb in length), which is flanked with short 5′ and 3′ non-coding regions (NCR) **[4]**. The genome comprises 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 **[5]**. Recent studies have determined the role of different reading frame encoded proteins in HEV regulation by analyzing their intrinsically disordered regions **[6 - 9]**, as these regions are linked with virus's infection and pathogenesis **[10 - 12]**. The proteins or protein regions that fail to get folded into definite three-dimensional (3D) structures but remain biologically active are termed as intrinsically disordered proteins (IDPs) and intrinsically disordered protein regions (IDPRs), respectively **[13 - 15]**. These disordered protein regions exist as extremely active ensembles that are rapidly interconvertible under different physiological conditions **[13 - 15]**. Due to occurrence of peculiar phenomenon, i.e., binding of several disordered regions to one ligand or vice-versa (one disordered region binds to many partners), the intrinsic disordered regions are utilized in protein-protein interactions **[16, 17]**. Therefore, it is of interest to document the intrinsically disordered regions in the rodent hepevirus proteome.

Materials and methods:

Sequence retrieval:

The protein sequence (Accession ID: GU345042) of rat HEV was obtained from the NCBI (National Center for Biotechnology Information) GenBank database.

Structural analysis:

The three-dimensional (3D) model of the rat HEV proteins was modeled using Phyre2 (Protein Homology/AnalogY Recognition Engine) server (http://www.sbg.bio.ic.ac.uk/~phyre2/html /page.cgi ?id=index) and analyzed.

Intrinsic disorder prediction:

Intrinsically disordered regions (IDRs) of the rat HEV proteome were predicted using the PONDR® (Predictor of Natural Disordered Regions) (www.pondr.com) at its default settings. Multiple predictors such as members of the PONDR® family including PONDR®VLS2, PONDR®VL3 and PONDR® VLXT were exploited to predict the intrinsic disorder predisposition in rat HEVs. This bioinformatics tool predicts the residues or regions which fail in propensity for an ordered structure formation. The protein residues with predicted scores between 0.2 and 0.5 were considered as flexible, while the residues which had scores, exceeding the 0.5 threshold value, were predicted as intrinsically disordered ones.

Results and Discussion:

Intrinsic disorder is linked with the pathogenesis and infection of the viruses **[10 - 12]**. To complete the life cycle, viruses require various interactions with the components of the host cells. Beginning from the virus's attachment, its entry, commandeering the host machinery, synthesis of the viral components and particle assembly to the last phase, i.e., exiting as new infectious particles from the host cell **[18]**. All these stages rely heavily on the intrinsic disorder prevalent in viral proteins **[18]**. The biology of the unstructured regions of the Norway rat HEV, comprising additional reading frames (ORF1, ORF2, ORF3, ORF4, ORF5 and ORF6) remains to be explored. Therefore, the present study reports the analysis on the unstructured regions of the ORF encoded proteins of rat hepevirus to shed novel light on its functionality in HEV regulation. Analysis of protein structure provides a detailed understanding of its function. Thus, the rat HEV protein structures were examined using a web portal for protein modeling and analysis. The modeled 3D structures of rat HEV proteins showed all the three secondary structure states, i.e., alpha helix, beta strand and loops/coils. A study has suggested that loops/coils are not necessarily disordered, however protein disorder is only found within loops **[19]**. The predicted percentage of disordered residues in the generated 3D rat HEV proteins are summarized in **Table 1**.

Table 1 Predicted disordered percentage in rat HEV proteome

Protein gene	GenBank Protein_ID	Secondary structure and disorder prediction
ORF1	ADM35748.1	Disordered (26%) Alpha helix (35%) Beta strand (16%)
ORF2	ADM35750.1	Disordered (42%) Alpha helix (18%)

		Beta strand (34%)
ORF3	ADM35749.1	Disordered (44%) Alpha helix (12%) Beta strand (12%)
ORF4	ADM35751.1	Disordered (16%) Alpha helix (20%) Beta strand (45%)
ORF5	ADM35752.1	Disordered (33%) Alpha helix (13%) Beta strand (34%)
ORF6	ADM35753.1	Disordered (22%) Alpha helix (30%) Beta strand (33%)

Therefore, the initial structural analysis revealed that all the rat HEV proteins consisted of disordered regions **(Figure 1A - F)**. The specific role of disordered regions in several nonstructural proteins has been demonstrated to participate in the multiplication and regulatory functions of viruses **[20]**. For instance, a recent study has shown the involvement of ORF4 protein in the regulation and pathogenesis of HEV due to the presence of significant fraction of disordered regions **[21]**. The disordered regions in the ORF1 Y-domain of HEV have also been shown to perform crucial role in its pathogenesis due to its intrinsic disorder phenomenon **[22]**. In HDV (hepatitis delta virus), the translation of a delta antigen (a single basic protein) forms the basis of its replication, which is considered as an IDP molecule **[23]**, via both experimental and computational studies **[24]**. The HCV (Hepatitis C virus) interacts with several viral and host proteins required for its replication via its disordered nonstructural NS5A protein domain **[25, 26]**. These protein–protein interactions result in the occurrence of several significant biological functions. Moreover, the PPR (Polyproline region) of nonstructural ORF1 has been associated with the regulation of HEV in addition to its role in replication, due to its characteristic intrinsic disorder property **[27]**. After the initial structural exploration, the intrinsic disorder propensity analysis of rat HEV proteins was carried out to elucidate their intrinsic disorder properties. The predicted intrinsic disordered residues obtained from three disorder predictors for ORF encoded proteins of rat HEV are mentioned in **Table 2.** The resulting disorder profiles of the rat HEV proteins are shown in **Figure 2A - F**.

The rat HEV proteins were initially categorized on the basis of overall degree of intrinsic disorder. The categories included structured proteins (0 - 10%), moderately disordered proteins (10 - 30%) and highly disordered proteins (30 - 100%) **[28-29]**. Additionally, the ORF proteins were categorized on the basis of length of disordered domains and overall fraction of disordered residues. The categories consisted of ordered proteins (ORDPs); intrinsically disordered protein regions (IDPRs) and intrinsically disordered proteins (IDPs) **[28- 30]**. ORDPs are intrinsic disorder protein variants which consist of less than 30% of disordered residues without disordered domain (consecutive disordered residues) at either C- or N-terminus or in positions distinct from the N- and C-terminals. IDPRs are intrinsic disorder protein variants which consist of less than 30% of disordered residues with disordered domain at either C- or N-terminus or in positions distinct from the N- and C-terminals. IDPs are intrinsic disorder protein variants which consist of more than 30% of disordered residues. Thus, based on these criteria, the obtained disorder profiles of different rat HEV ORF proteins from different disorder predictors are discussed below.

Table 2: The predicted percentage of intrinsic disorder scores of ORF proteins of rat HEV

Predictor	Disordered regions	Overall percent disordered	Number of disordered regions	Longest disordered region	Average Prediction score
ORF1					
VLXT	[12-18], [87-92], [110-133], [232-245], [275-280], [501-508], [576-633], [643-643], [651-676], [680-683], [746-758], [801-805], [830-892], [1013-1018], [1054-1059], [1158-1193], [1203-1210], [1269-1271], [1288-1306], [1454-1458, [1463-1463], [1557-1576], [1635-1636]	20.90	23	63	0.2824
VL3	[112-130], [486-507], [576-577], [827-888], [1172-1199]	14.24	5	102	0.2727
VSL2	[1-8], [11-22], [109-122], [173-176], [225-235], [571-675], [712-714], [762-765], [823-887], [894-897], [926-931], [966-967], [1089-1091], [1164-1199], [1410-1415], [1456-1466], [1561-1572], [1631-1636]	19.50	18	105	0.3395
ORF2					
VLXT	[25-55], [73-168], [211-230], [248-266], [312-323], [385-406], [441-454], [467-480], [580-607], [637-644]	40.99	10	96	0.4149
VL3	[27-138]	17.39	1	112	0.2934
VSL2	[25-143], [209-214], [250-261], [304-310], [319-323], [390-399], [432-451], [470-480], [515-518], [594-603], [636-644]	33.07	11	119	0.4360
ORF3					
VLXT	[29-102]	72.55	1	74	0.7093

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 18(2): 111-118 (2022)

VL3	[1-102]	100.00	1	102	0.9208
VSL2	[1-102]	100.00	1	102	0.9475
ORF4					
VLXT	[1-28], [41-41], [76-81], [117-118]	20.22	4	28	0.2747
VL3	[1-38]	20.77	1	38	0.3077
VSL2	[1-47], [118-121]	27.87	2	47	0.3321
ORF5					
VLXT	[20-27], [51-52], [70-155], [159-178], [195-201]	60.00	5	86	0.6284
VL3	[65-180]	56.59	1	116	0.5566
VSL2	[42-42], [63-182], [200-205]	61.95	3	120	0.6423
ORF6					
VLXT	[41-47], [81-91]	19.78	2	11	0.2359
VL3	-	0.00	0	0	0.2278
VSL2	[1-3], [29-33], [84-91]	17.58	3	8	0.2552

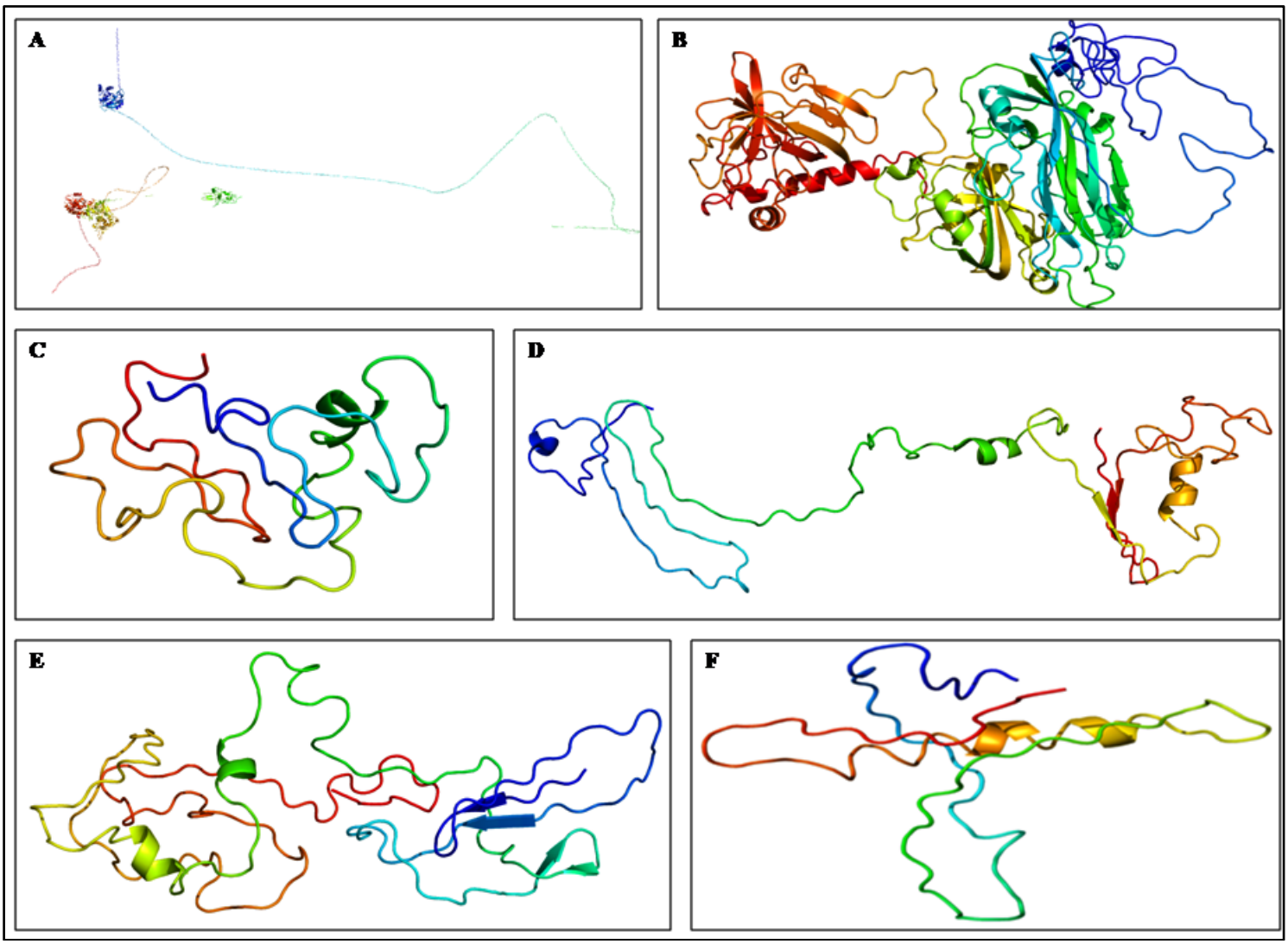

Figure 1: Representation of the modeled three dimensional structures of ORF proteins of rat HEV. (A) ORF1; (B) ORF2; (C) ORF3; (D) ORF4; (E) ORF5; and (F) ORF6. The analysis was conducted using Phyre2 webserver. The images are coloured by rainbow N → C terminus.

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 18(2): 111-118 (2022)

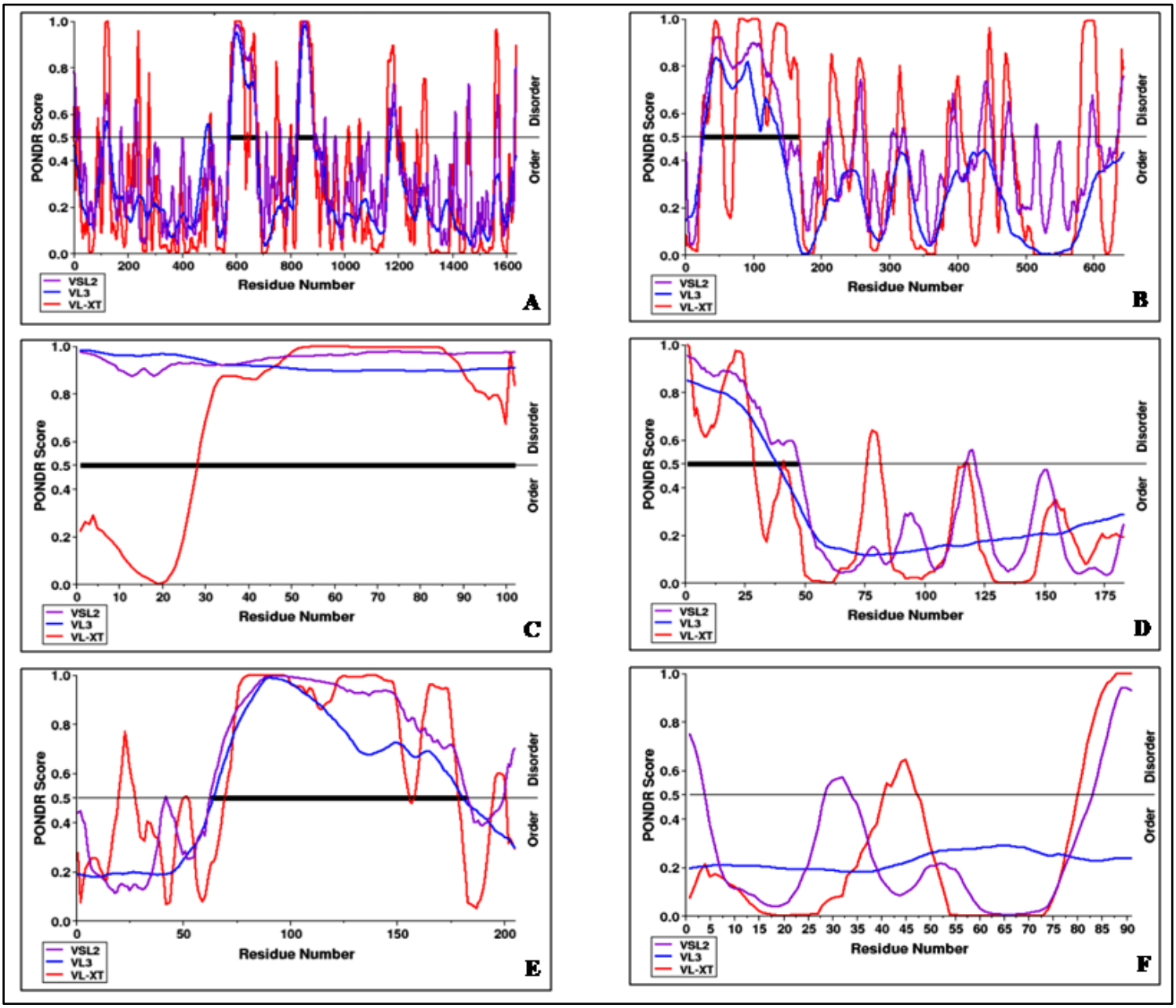

Figure 2: Analysis of intrinsic disorder predisposition of rat HEV proteome. Figure 1A - F represents the intrinsic disorder profiles of rat HEV proteins: (A) ORF1; (B) ORF2; (C) ORF3; (D) ORF4; (E) ORF5; and (F) ORF6. Disorder probability was calculated using three different predictors of PONDR family, i.e., PONDR VSL2, PONDR VL3 and PONDR VL-XT, 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.

ORF1: The intrinsic disorder profile of the ORF1 protein showed it as a moderately disordered protein, as it consisted of less than 30% (VLXT, VSL2 and VL3) of the disordered residues in its polypeptide chain with two stretches of disordered domains at positions distinct from N- and C-terminals. Thus, were categorized into IDPRs, i.e., structured proteins with intrinsically disordered segments or proteins possessing both structured unstructured regions **(Table 2)**.

ORF2: The intrinsic disorder profile of the ORF2 protein showed it as a highly disordered protein, as it consisted of >30% (as predicted by 40.99% VLXT and 33.07% VSL2) and moderately disordered (17.39% by VL3) as it consisted of less than 30% of the disordered residues in its polypeptide chain along with presence of disordered domain. Thus, on combining these assumptions it was categorized into both IDPs, i.e., proteins having significant fraction of

disordered regions or IDPRs, i.e., structured proteins with intrinsically disordered segments **(Table 2).**

ORF3: The intrinsic disorder profile of the ORF3 protein showed it as a highly disordered protein, as it consisted of >30% (VLXT, VSL2 and VL3) of the disordered residues in its polypeptide chain. Thus, were categorized into IDPs **(Table 2).**

ORF4: The intrinsic disorder profile of the ORF4 protein showed it as a moderately disordered protein, as it consisted of less than 30% (VLXT, VSL2 and VL3) of the disordered residues in its polypeptide chain with a stretch of disordered domain at N-terminus. Thus, were categorized into IDPRs, i.e., structured proteins with intrinsically disordered segments **(Table 2)**.

ORF5: The intrinsic disorder profile of the ORF3 protein showed it as a highly disordered protein, as it consisted of >30% (VLXT, VSL2 and VL3) of the disordered residues in its polypeptide chain along with possession of long disordered domain towards C-terminus. Thus, were categorized into IDPs **(Table 2)**.

ORF6: The intrinsic disorder profile of the ORF6 protein showed it as a structured protein, as it consisted of less than 30% (VLXT, VSL2 and VL3) of the disordered residues in its polypeptide chain without the presence of any disordered domain. Thus, were categorized into ORDPs, i.e., proteins possessing significant amount of structure **(Table 2)**.

Table 3 Categorization of rat HEV proteins based on intrinsic disorder variant

Protein	Overall degree of intrinsic disorder	On basis of disordered domain length and overall fraction of disordered residues
ORF1	Moderately disordered	IDPR
ORF2	Moderately/highly disordered	IDPR/IDP
ORF3	Highly disordered	IDP
ORF4	Moderately disordered	IDPR
ORF5	Highly disordered	IDP
ORF6	Structured	ORDP

Thus, our intrinsic disorder propensity analysis revealed ORF6 as a highly structured protein (ORDP); ORF1 and ORF4 as moderately disordered proteins (IDPRs); ORF3 and ORF5 as highly disordered proteins (IDPs); and ORF2 as both moderately disordered protein (IPPR) or highly disordered protein (IDP) **(Table 3)**. Interestingly, the computational analysis of revealed ORF3 protein as the most disordered protein, while ORF6 protein as the most ordered protein in the rat HEV proteome with the remaining proteins as intermediates. In context with this, it is noteworthy to mention that all the three intrinsic disorder variants were found in the rat HEV proteins, i.e., ORDP, IDPR and IDP **(Table 3)**. The "IDPR/IDP" is defined as disordered region in protein or disordered protein. These regions/proteins perform a number of significant roles in a variety of biological processes, such as control of signaling pathways, cell cycle regulation, etc. **[11, 16, 17, and 31]**. It has been suggested that IDPRs/IDPs achieve their signaling cascade by binding to their partners with low affinity and high specificity **[32]**. Thus, the proteins, such as ORF1, ORF2 and ORF4 can play crucial roles in important biological processes as IDPRs. IDP plays significant role in recognition, signaling, regulation and control of protein-protein interaction (PPI) networks **[33]**. IDPs form essential components of cellular signaling machinery due to its ability to interact differently which results in different consequences **[34]**. Moreover, they are characterized by enormous flexibility and random conformation (coiled-like). Thus, these distinctive features enable IDPs to participate in one to many and vice-versa interaction **[35]**.

Conclusion:

The occurrence of the unstructured regions in rat HEV protein sequences suggested their disorder-based binding tendency. The intrinsic disorder analysis revealed all the three intrinsic disorder variants in the proteome of rat HEV. Further, the ORF3 protein was identified as the most disordered protein and ORF6 protein consisted of least fraction of intrinsic disorder. Thus, the rat HEV proteins could be engaged in diverse and essential biological functions.

Acknowledgement:

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.

Funding:

Not applicable

References: [Please give PMID or DOI]

[1] Kumar S *et al. Int J Infect Dis.* 2013 **17:**e228 [PMID: 23313154].
[2] Khuroo MS & Khuroo MS, *J Viral Hepat.* 2016 **23:**68. [PMID: 26344932]
[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 XJ, *Cold Spring Harb Perspect Med.* 2019 **9:**a031724 [PMID: 29530948].
[6] Shafat Z *et al. Netw Biol.* 2021 **11:**295. [PMID or DOI]
[7] Shafat Z *et al. J Genet Eng Biotechnol.* 2021 12**:**e1005521 [PMID: 34637041].
[8] Shafat Z *et al. Protein Expr Purif.* 2021 **187:**105947 [PMID: 34314826].
[9] Shafat Z *et al. Bioinformatics* 2021 **17:**818. [PMID or DOI]
[10] Singh A *et al. Sci Rep.* 2018 **8:**582 [PMID: 29643398].

[11] Giri R *et al. Front Cell Infect Microbiol.* 2016 **6**:144 [PMID: 27867910].
[12] Xue B *et al. Protein Pept Lett.* 2010 **17:**932 [PMID: 20450483].
[13] Lee R *et al. Chem Rev.* 2014 **114:**6589 [PMID: 24773235].
[14] Oldfeld CJ & Dunker AK, *Annu Rev Biochem*. 2014 **83:**553. [PMID: 24606139]
[15] Wright PE & Dyson HJ, *J Mol Biol*.1999 **293:**321 [PMID: 10550212].
[16] Dyson HJ & Wright PE, *Nat Rev Mol Cell Biol.* 2005 **6:**197 {PMID: 15738986].
[17] Dyson HJ & Wright PE, *Curr Opin Struct Biol*. 2002 **12:**54 [PMID: 11839490].
[18] Xue B *et al. Chem Rev.* 2014 **114:**6880 [PMID: 24823319].
[19] Linding R *et al. Structure* 2003 **11:**1453 [PMID: 14604535]
[20] Mishra PM *et al. Prog Mol Biol Transl Sci.* 2020 **174**:1 [PMID: 32828463].
[21] Shafat Z *et al. J Proteins Proteom.* 2021 **12:**289. [PMID or DOI]
[22] Shafat Z *et al. J Genet Eng Biotechnol.* 2021 **19:**154 [PMID: 34637041].
[23] Rizzetto M *J Hepatol*. 2009 **50:**1043 [PMID: 19285743].
[24] Alves C *et al. Virology* 2010 **407:**333 [PMID: 28055099].
[25] Verdegem D *et al. J Biol Chem.* 2011 [PMID: 21489988].
[26] Macdonald A & Harris M, *J Gen Virol* 2004 **85:**2485 [PMID: 15302943].
[27] Purdy MA *et al. PLoS ONE* 2012 **7:**e35974 [PMID: 22545153].
[28] Gsponer J *et al. Science* 2008 **322:**1365 [PMID: 19039133].
[29] Edwards YJ *et al*. *Genome Biol.* **10:**R50 [PMID: 19432952].
[30] Lee R *et al. Cell Rep*. 2014 **8:**1832 [PMID: 25220455].
[31] Gadhave K *et al. Cell Mol Life Sci*. 2020 **77:**4163 [PMID: 31894361].
[32] Uversky VN *Int J Biochem Cell Biol*. 2011 **43:**1090 [PMID: 21501695]
[33] Dafforn TR & Smith CJ EMBO Rep 2004 **5:**1046 [PMID: 15520805].
[34] Wright PE & Dyson HJ, *Nat Rev Mol Cell Biol.* 2015 **16:**18 [PMID: 25531225].
[35] Wang J *et al. Int J Mol Sci.*2011 **12:**3205 [PMID: 21686180].

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 18(2): 111-118 (2022)