Article

Shedding light on the dark proteome of Hepatitis E Virus

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: zoyashafat26@gmail.com, anahmed@ksu.edu.sa, mohkhalid@ksu.edu.sa, sparveen2@jmi.aci.in

Received 11 August 2021; *Accepted 18 September 2021*; *Published 1 December 2021*

Abstract

Hepatitis E virus (HEV) is a quasi-enveloped RNA virus of the family *Hepeviridae*. HEV is the chief cause of acute hepatitis worldwide, causing approximately 20 million infections annually, which results in 60,000 deaths. Due to insufficiency in appropriate HEV in vitro cell culture systems, our knowledge of its pathogenesis is inadequately understood. HEV encodes three open reading frames (ORFs): ORF1 (replicative machinery), ORF2 (viral capsid) essential for (infectious particles formation) and ORF3 (viral release). The presence of known and unknown coding and non-coding regions of HEV ORFs are still debated. Viral proteins entail disordered regions which are linked with the infectivity and pathogenicity of virus. Thus, we examined the dark proteome of HEV through analyzing intrinsically disordered protein regions (IDPRs) present in the ORFs by exploiting computational methodologies. Our findings suggested that ORF3 had the highest prevalence of disordered regions. The ORF3 region was followed by ORF2, which had comparatively lesser fraction of intrinsic disorder. The ORF1 had the least number of disordered residues in the HEV proteome. Our intrinsic disorder analysis results revealed that ORF1 polyprotein consists of mostly ordered domains, i.e., proteins having significant level of well-defined structures, with the exclusion of Pro and PCP domains. The analysis reveals Pro domain as a highly disordered protein while PCP domain as an intrinsic disordered protein. MoRF analysis revealed that HEV proteome contains multiple MoRFs across all ORFs. IDPRs are characterized by remarkable conformational flexibility and structural plasticity resulting in their engagement in several biological processes. Due to possession of MoRF in the HEV proteins, these regions can be used for protein–protein interactions due to the structural flexibility. These extensive findings on the HEV proteome will have significant implications in understanding the deeper functioning of structural as well as non-structural biology of HEV proteins.

Keywords HEV; ORF1; ORF2; ORF3; intrinsically disordered proteins; protein-binding propensity; DNA-binding propensity; RNA-binding propensity.

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 a quasi-enveloped RNA virus of the family *Hepeviridae*. The HEV genome is systematized into three partially over-lapped ORFs (ORF1, ORF2 and ORF3) (Tam et al., 1991). The largest ORF1 encodes a non-structural polyprotein, required for the replication of HEV (Ansari et al., 2000; Parvez, 2013). The non-structural polyprotein consists of multiple functional domains including methyltransferase (Met: 56 - 240 aa), Y (Y: 216 - 442 aa), putative papain-like cysteine protease (PCP: 433 - 592 aa), hypervariable region (HVR: 593 - 711 aa), proline region (Pro: 712 - 778 aa), X (X: 785 - 942), RNA helicase (Hel: 960 - 1204 aa), and the RNA-dependent RNA polymerase (RdRp: 1207 - 1693 aa) (Koonin et al., 1992; Nan and Zhang, 2016; Proudfoot et al., 2019). The domains such as PCP, X, Y, and HVR are still precisely understood (LeDesma et al., 2019). Molecular studies have reported the essential role played by various ORF1 domains which include Met (Kabrane-Lazizi et al., 1999; Magden et al., 2001; Zhang et al., 2001), RdRp (Agrawal et al., 2001; Mahilkar et al., 2016), HVR (Pudupakam et al., 2009; Pudupakam et al., 2011) and Hel (Karpe and Lole, 2010; Karpe and Lole, 2010) in HEV replication. Moreover, the undefined X (Parvez, 2015; Ojha and Lole, 2016; Anang et al., 2016) and Y (Parvez, 2017) domains have recently been reported to perform critical roles in the life cycle of HEV. However, direct correlations between the function of these putative protein domains and viral pathogenesis are still debated due to lack of information. The ORF2 encodes the viral capsid protein which forms the major structural component in HEV virions (Chandra et al., 2008; Mori and Matsuura, 2011). The ORF3 phosphoprotein is implicated in modulation of host cell signaling pathways, virus infectivity and virion release (He et al., 2016; Parvez and Al-Dosari, 2015; Ding et al., 2017).

The present study analyzed the structurally “unknown” regions (i.e., a fraction of a proteome that has no detectable similarity to any PDB structure) of the HEV proteome. This fraction we call it as the “dark proteome”. These regions were characterized to shed some light on interplay between the ordered and disordered cellular machineries using computational tools for ubiquity of the intrinsic disorder. The proteins or protein regions that fail to get folded into of definite three-dimensional (3D) structures but remains biologically active are termed as intrinsically disordered proteins (IDPs) and intrinsically disordered protein regions (IDPRs) respectively. These disordered protein regions exist as extremely active ensembles that are rapidly interconvertible under different physiological conditions (Van et al., 2014; Oldfeld and Dunker, 2014; Wright and Dyson, 1999). Intrinsically disordered proteins (IDPs) are involved in several biological processes (cellular signalling, gene regulation, etc.) enable them to perform various biological functions by interacting with their physiological partners (Dunker et al., 2005; Dunker et al., 2002; Dunker et al., 2008; Liu et al., 2006; Uversky et al., 2005). The IDPs and IDPRs functions are controlled by their protein–protein, protein–RNA, and protein–DNA interactions (Yan and Kurgan, 2017; Peng, 2015). Molecular recognition features (MoRFs) are the regions in IDPs implicated in the regulation of its function by protein–protein interactions and serve as the primary stage in molecular recognition.

In viruses, the IDPs/IDPRs are often strongly associated with virulence as they encompass unstructured regions (Giri et al., 2016; Xue et al., 2010; Singh et al., 2018; Ward et al., 2004). Moreover, protein-protein interaction is a potential source for drug targets (Midic et al., 2009). 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 (Uversky et al., 2005; Dunker et al., 1998). Thus, the intrinsic disordered region in proteins are considered as potential drug targets due to disordered to ordered transition state upon drug binding (Salma et al., 2009). The current study reports analysis on disordered side of HEV proteome using combination of different computational methods to check the occurrence of IDPRs in order to gain insights into their disorder-related functions.

Further, the identification of protein functions in terms of protein–protein interactions, RNA binding, and DNA binding were also investigated to shed some light on the sequence and structural idiosyncrasies of HEV evolution.

2 Material and Methods

2.1 Sequence retrieval

The protein sequence of HEV was retrieved from the GenBank database (Accession ID: NC_001434) (Clark et al., 2016). This reference sequence was used for multiple predictions.

2.2 Predictions of intrinsic disorder predisposition

The online predictor DisProt (version of PONDR-FIT) was used to predict the intrinsic disorder predisposition in HEV proteome (Xue et al., 2010). We used this tool over other members of the PONDR® (Predictor of Natural Disordered Regions) family (Peng et al., 2005) as it is slightly better than PONDR®VSL2 (Peng et al., 2006), when both ordered and disordered protein regions occur in the sequences. 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.

2.3 Identification of protein-binding regions

MoRFpred (Disfani et al., 2012; Metallo, 2010) online bioinformatics predictor was used for the prediction of Molecular Recognition Feature (MoRF) in HEV protein sequences. The residues which scored above the threshold value of 0.5 were considered as MoRF regions. MoRFpred predicts residues which have high propensity to bind to protein partners.

2.4 Identification of DNA-binding regions

DRNAPred (Yan and Kurgan, 2017) and DisoRDPbind (Peng et al., 2017) online servers were used for the prediction of DNA-binding residues in HEV proteome.

2.5 Identification of RNA-binding regions

PPRInt (Prediction of Protein RNA-Interaction) (Kumar et al., 2008) and DisoRDPbind (Peng et al., 2017) servers were used for the prediction of RNA-binding residues in the HEV protein sequences.

3 Results

The IDPs/IDPRs prevalence in viral proteomes emerged as a novel concept and was defined as “prevalence of exclusivity/exceptionality” concept by researchers (Uversky, 2015). The HEV genome is systematized into three ORFs (ORF1, ORF2 and ORF3) (Fig. 1) (Parvez, 2017). The predicted percentage of disordered residues obtained from Phyre2 webserver for HEV ORFs are summarized in Table 1 and Supplementary Figs S1, S2, and S3. The predicted 3D structures of the HEV proteins are shown in Supplementary Fig. S4.

3.1 Analysis of ORF1 non-structural domains

ORF1: The non-structural polyprotein consists of eight domains: Methyltransferase (Met: 56 - 240 aa), Y (Y: 216 - 442 aa); Putative papain-like cysteine protease (PCP: 433 - 592 aa); Hypervariable region (HVR: 593 - 711 aa); Proline region (Pro: 712 - 778 aa); X (X: 785 - 942); RNA helicase (Hel: 960 - 1204 aa); and RNA-dependent RNA polymerase (RdRp: 1207 - 1693 aa) (Purdy, 2012).

Table 1 Predicted disordered regions in HEV proteins.

Protein genes	**Length of the protein** (NCBI Reference Sequence accession ID)	**Secondary structure and disorder prediction**
ORF1	NP_056779.1	Disordered (19%) Alpha helix (35%) Beta strand (22%)
Met domain (56 - 240 aa)	NP_056779.1	Disordered (17%) Alpha helix (36%) Beta strand (26%)
Ydomain (216 - 442 aa)	NP_056779.1	Disordered (4%) Alpha helix (55%) Beta strand (13%)
PCP domain (433 - 592 aa)	NP_056779.1	Disordered (31%) Alpha helix (24%) Beta strand (36%)
HVR domain (593 - 711 aa)	NP_056779.1	Disordered (25%) Alpha helix (24%) Beta strand (32%)
Proline domain (712 - 778 aa)	NP_056779.1	Disordered (93%) Alpha helix (16%) Beta strand (0%)
Xdomain (785 - 942 aa)	NP_056779.1	Disordered (14%) Alpha helix (35%) Beta strand (17%)
Hel domain (960 - 1204 aa)	NP_056779.1	Disordered (16%) Alpha helix (25%) Beta strand (32%)
RdRp domain (1207 - 1693 aa)	NP_056779.1	Disordered (12%) Alpha helix (51%) Beta strand (11%)
ORF2	NP_056788.1	Disordered (50%) Alpha helix (13%) Beta strand (33%)
ORF3	YP_003864075.1	Disordered (18%) Alpha helix (25%) Beta strand (19%)

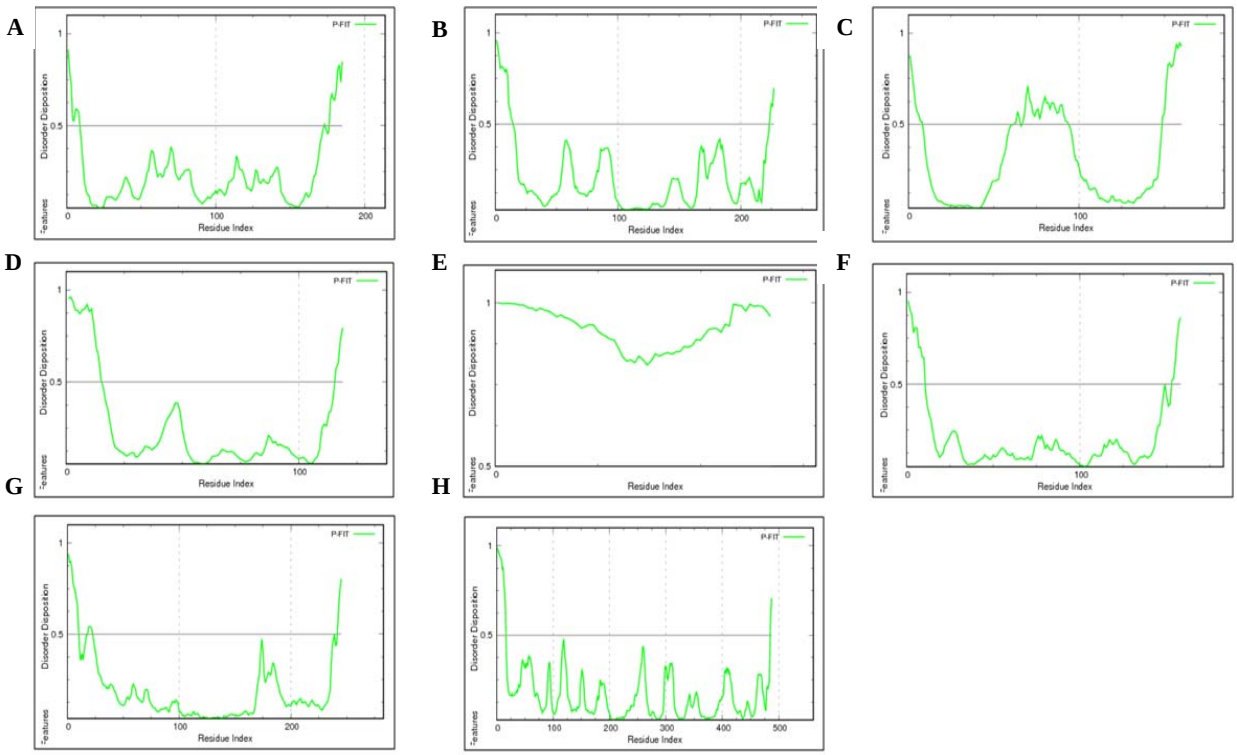

Fig. 1 Analysis of intrinsic disorder predisposition of HEV proteome. Figure 1A - H represents the intrinsic disorder profiles of HEV proteins: (A) Met domain; (B) Y-domain; (C) PCP domain; (D) HVR domain; (E) Pro domain; (F) X-domain; (G) Hel domain; and (H) RdRp domain. Disorder probability was calculated using DisProt, 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.

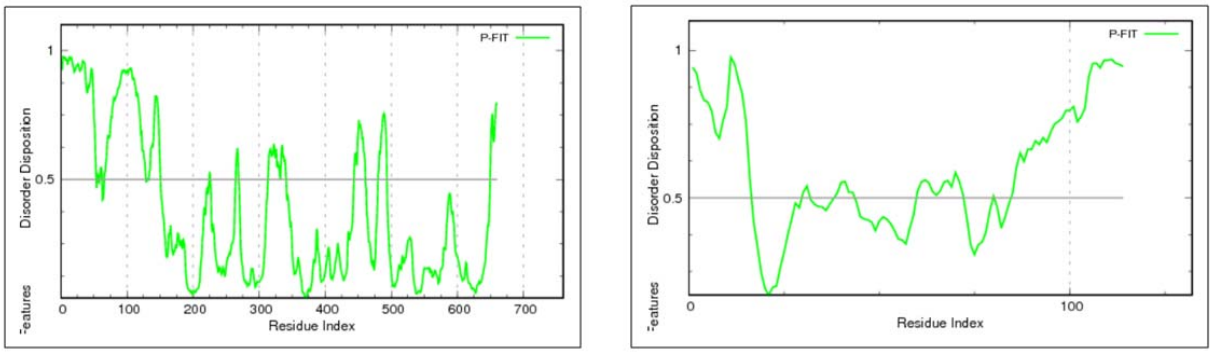

Fig. 2 Analysis of intrinsic disorder predisposition of HEV proteome. Representation of intrinsic disorder profile in HEV proteins: (A) ORF2 and (B) ORF3. Disorder probability was calculated using DisProt, where graphs represent the intrinsic disorder profiles. A threshold value of 0.5 was set to distinguish between ordered and disordered region along the genome (dashed line). Regions above the threshold are predicted to be disordered.

Fig. 3 Analysis of protein-binding propensity of HEV proteome. The resulting protein-binding profile was calculated using MoRFPred. The protein binding residues are depicted in red.

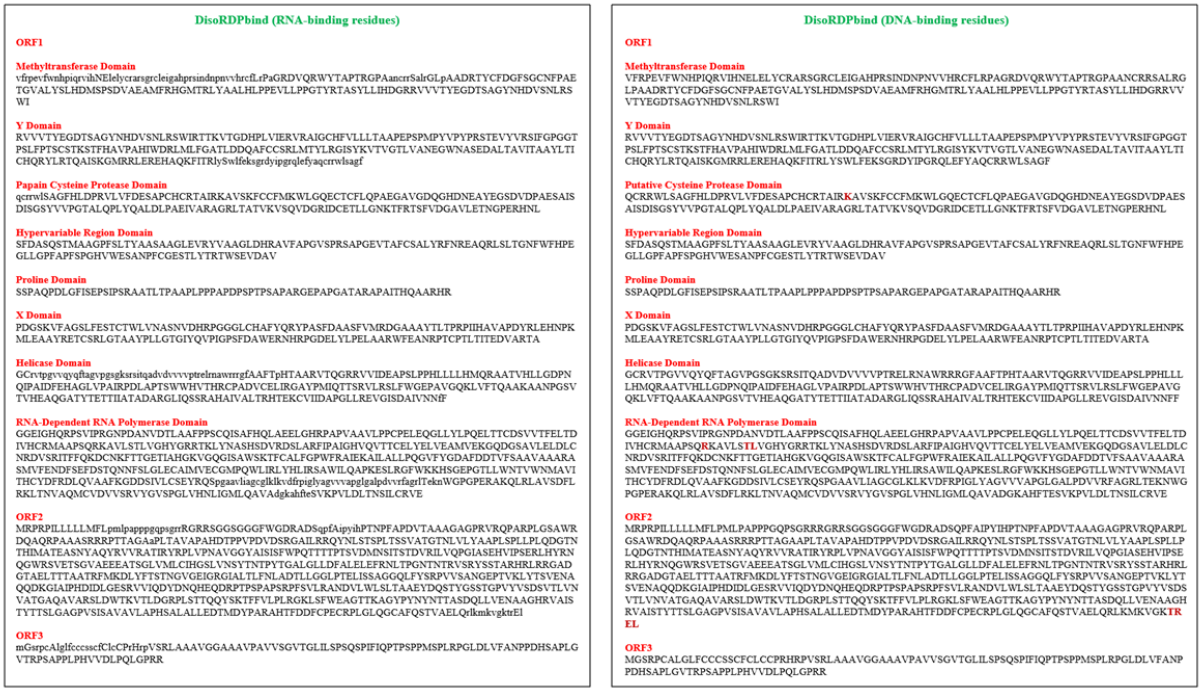

Fig. 4 Analysis of nucleotide-binding propensity of HEV proteome. The left panel shows the resulting RNA-binding profile and the right panel shows the DNA-binding profile. The nucleotide-binding propensity for HEV proteins was calculated using DisoRDPbind. The identified nucleotide-binding residues are depicted in red.

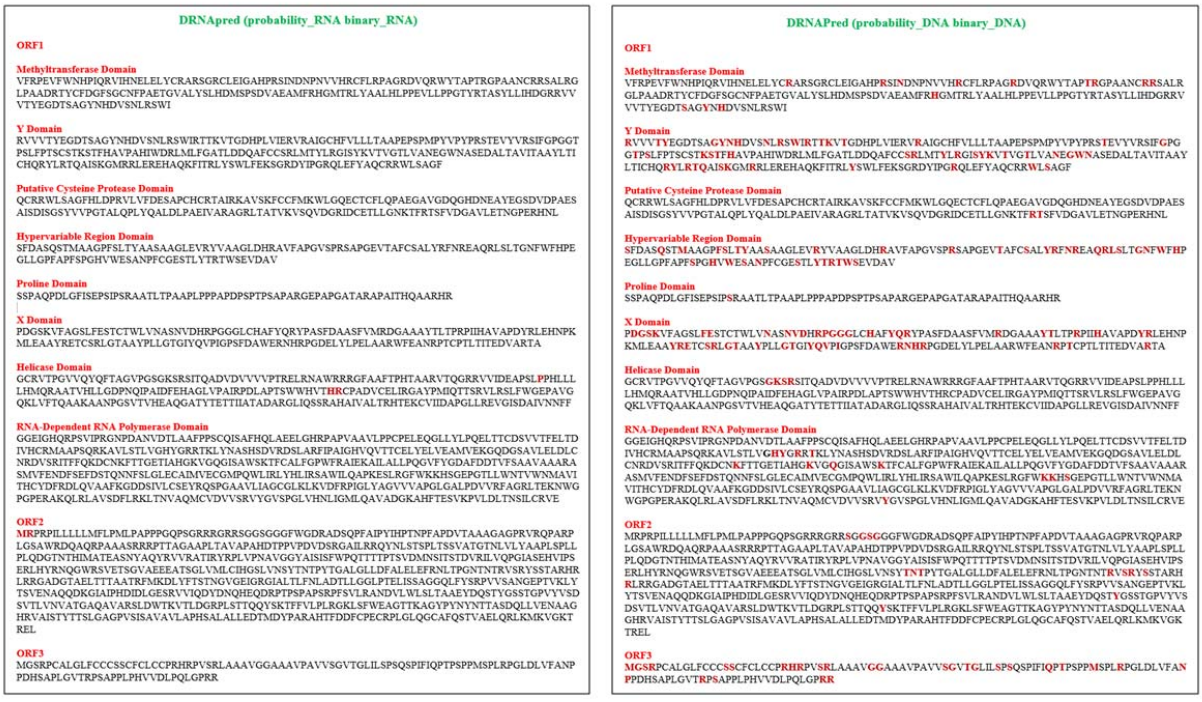

Fig. 5 Analysis of nucleotide-binding propensity of HEV proteome. The left panel shows the resulting RNA-binding profile and the right panel shows the DNA-binding profile. The resulting nucleotide-binding profile was calculated using DRNApred. The identified nucleotide-binding residues are depicted in red.

Fig. 6 Analysis of RNA-binding propensity of HEV proteome. The resulting RNA-binding profile was calculated using PPRInt webserver. The identified RNA-binding residues are depicted in red.

3.1.1 Met domain

Based on the sequence homology, the ORF1 region spanning 60 - 240 aa residues suggested as putative methyltransferase (Koonin et al., 1992). Literature evidences have showed discrepancies in determining the endpoint of Met domain (Koonin et al., 1992). The intrinsic disorder profile of Met domain of HEV obtained from disorder predictor showed that Met protein consists of an ordered structure (Fig. 1A) with highly flexible N and C-terminals. The protein possessed a MoRF region only at its C-terminal (residues 1265 - 1272) as predicted by MoRFPred, which clearly suggests that disordered residues play significant interactive role in the Met activity (Fig. 3). Besides this, presence of MoRF further demonstrates its interaction with the Y-domain protein for the processing of ORF1 polyprotein. Further, Few RNA- and DNA-nucleotide binding residues were predicted by DisoRDPbind (Fig. 4) and DRNAPred (Fig. 5) respectively assisting in interactions. Several mutations in the Met domain have been suggested to play significant role in patients experiencing acute liver failure (ALF) (D29N and V27A) and viremia (H105R). Thus, this region in future could serve as a potential therapeutic target (Borkakoti et al., 2017).

3.1.2 Y-domain

The disorder profile of Y-domain obtained from disorder predictor showed it to be a structured protein however the protein possessed multiple flexible regions (Fig. 1B). The protein consisted of a MoRF region only at its C-terminal as predicted by MoRFPred (Fig. 3). This demonstrates its interaction with the PCP protein for the processing of ORF1 polyprotein. Further, several RNA and DNA-binding sites were predicted by DisoRDPbind (Fig. 4), DRNApred (Fig. 5) and PPRInt (Fig. 6) which suggests strong tendency of the Y-domain to bind to RNA as well as DNA. Furthermore, a highly conserved α-helix counterpart 'LYSWLFE' (aa 410–416) has been predicted in the Y-domain, required for its cytoplasmic membrane binding (Parvez, 2017). Presence of MoRF region in this conserved counterpart clearly suggests the Y-domain essentiality in binding. Thus, our results support the previous literature (Parvez, 2017). Also, substitutions of the universally conserved residues (L410, S412 and W413) in the conserved α-helix motif have been reported to abolish RNA replication (Parvez, 2017). Several RNA-binding residues at the C-terminus (including conserved counterpart) clearly indicate its essentiality in binding with the viral RNA. Further, several DNA-binding (as predicted by DRNAPred) residues showed high propensity for this protein to bind to DNA. Thus, our results show consistency with previous reports which demonstrate the essentiality of Y-domain in HEV (Parvez, 2017; Shafat et al., 2021). Though these findings enhance our knowledge on this precisely understood Y-domain, however further information is still required to delineate its function and its conserved residues criticality in the viral replication.

3.1.3 PCP domain

The intrinsic disorder analysis revealed PCP domain as a protein having a long-disordered region (>30 amino acid residues), thus suggesting it as an IDPR (intrinsic disordered protein region). The disorder profile of PCP domain obtained from the disorder predictor has been shown (Fig. 1C). The PCP protein consisted of only single MoRF at its N-terminal as predicted by MoRFPred (Fig. 3). Furthermore, only few RNA and DNA-binding sites were identified by DisoRDPbind (Fig. 4) and DRNApred webservers respectively (Fig. 5). The PCP region was initially recognized as a putative protease due to its sequence homology with the rubella virus (Koonin et al., 1992). Studies on PCP in various expression models have beeninspected, leading to contradicting results. The two major paradigms regarding this domain state: (1) PCP process the HEV ORF1 polyprotein into distinct functional units (Parvez and Khan, 2014; Sehgal et al., 2006; Parvez, 2013; Paliwal et al., 2014); and (2) PCP does not process the ORF1 polyprotein, and in order to perform proteolytic function, their targets should be cellular instead of viral (Kanade et al., 2018). The putative protease region of HEV and its function(s) are still under debate.

3.1.4 X domain

X domain is also referred to as 'macro domain', as it resembles non-histone domain of the histone macroH2A (Parvez, 2015).The disorder profile of the X-domain shows it to have an ordered structure with possession of residues at N-terminus displaying higher tendency for intrinsic disorder (Fig. 1D). MoRFPred analysis revealed that the protein consisted of a single MoRF region at its N-terminus (Fig. 3). Further, several DNA-binding (Fig. 5) and few RNA-binding sites (Fig. 6) were identified in the polypeptide chain (DRNApred and PPRInt). The prediction of several RNA-binding residues in the potential Appr1″-pase active site (Asn806, Asn809, His812, Gly815, Gly816 and Gly817) at N-terminus supports the earlier study, which suggests significant role played by X-domain in the HEV replication (Parvez, 2015). Further, the identified propensity of the 'NxxNxxHxxGGG' segment towards both RNA and DNA residues show consistency with the earlier observations suggesting catalytic/regulatory function performed by the non-structural X-domain (Parvez, 2015). However, complete information on the X-domain function is yet to be explored.

3.1.5 HVR and Pro domains

Between the PCP domain and the X domain, there are HVR and Pro domains. Initially, Tsarev and colleagues considered these two domains as a common entity and named it as HVR (712 - 778 aa) due to HEV Sar55 extreme sequence divergence when compared with other strains (Tsarev et al., 1992). But later report showed a separate region designated as a proline rich region (712 - 778 aa) (Dosztanyi et al., 2006; Dunker et al., 2008). However, recent investigations have designated the specific ORF1 region, i.e., 712 - 778 aa region as the HVR domain, which was formerly recognized as Pro domain, without considering the immediate upstream region (592 - 711 aa) (Pudupakam et al., 2011; Pudupakam et al., 2009). But on the other hand, reports also claim HVR and Pro as two distinct HEV ORF1 proteins: HVR domain (593 - 711 aa) and Proline-rich region (712 - 778 aa) (Purdy et al., 2012). Inconsistency still prevails in the nomenclature of HVR and Pro regions. Thus, we did perform the disorder analysis for these two regions separately.

HVR domain

The graph showing the disorder analysis of the HVR protein reveals it has a well-defined structure with high disorder propensity at N-terminus (short stretch of disordered region less than 30 aa residues) with slightly flexible C-terminus. MoRF analysis revealed the presence of MoRF regions (C-terminus) and few MoRF residues at the N-terminus (Fig. 1E). These results indicate that intrinsic disorder play a crucial role in the interactive ability of HVR region with its binding partners. The nucleotide-binding residues were predicted using different webservers (DisoRDPbind, DRNAPred and PPRInt), which showed the absence of HVR residues propensity in binding to RNA (Fig. 4, 5, 6). However, several DNA-binding residues were identified at both N- and C-terminus as well as along the entire HVR polypeptide chain by DRNAPred (Fig. 5). Since, most of the publications have mainly focused on the Pro region instead of the upstream HVR domain, thus its function remains to be elucidated.

Pro domain

The Pro domain serves as a hinge between its upstream and X domain domains due to possession of multiple proline residues in its polypeptide chain. Due to this peculiarity in the protein polypeptide, it lacks a stable tertiary structure (Purdy et al., 2012). The disorder graph profile shows that Pro domain has a complete disordered structure, suggesting it as an IDP (intrinsic disordered protein), supports the previous report (Fig. 1F) (Purdy et al., 2012). The protein consisted of MoRFs at both the terminals; N-terminus (three to seven) C-terminus (nine residues) as predicted by MoRFPred webserver (Fig. 3). Further, several protein-binding sites were also identified using the DisoRDPbind server (S5 Figure). Thus, the results obtained in the present study suggest disorder-to-order state transition in the Pro region, which are consistent with the earlier observations (Purdy et al., 2012). Further, the predicted RNA- and DNA-binding tendency in the proline region showed the absence of nucleotide-binding sites (Fig. 4, 5, 6), except one DNA-binding residue (as predicted by DRNAPred webserver) (Fig. 5). This clearly reveals the non-essentiality of the Pro domain in viral replication. This is coherent with the previous report which suggested that Pro domain is not required for the replication of virus and its infectivity, but performs a role in replication efficiency (Pudupakam et al., 2011; Pudupakam et al., 2009).

3.1.6 Hel domain

The domain is essential for the replication of positive-stranded RNA viruses (Kadare and Haenni, 1997). Helicases unwind the strands of nucleic acid by acting as motor proteins using energy from ATP hydrolysis (Kadare and Haenni, 1997). The intrinsic disorder analysis revealed that the Hel protein has an ordered structure, with flexible N-terminus but contain virtually no IDPRs (Fig. 1G). The MoRFs were predicted at both the N- and C-terminals, which suggest the protein-protein interactive ability (Fig. 3) (Karpe and Lole, 2010). The Hel domain consists of two conserved motifs: Walker A (GVPGSGKS; aa 975–982) and Walker

B (DEAP; aa 1029–1032), which has been demonstrated to participate in purine nucleoside triphosphate (NTP)-binding activity (Tam et al., 1991). Presence of nucleotide-binding residues in these conserved regions was observed. Mutations in this domain have been linked to cases of fulminant hepatic failure (Smith and Simmonds, 2015; Devhare et al., 2014). To sum up these results, it can be concluded that Hel region plays an essential role in the life cycle of HEV and mutation-associated disease severity.

3.1.7 RdRp domain

The RdRp protein found in RNA viruses (positive sense) is necessary for their genome replication (O' Reilly and Kao, 1998). This particular domain contains approximately 300 residues in its polypeptide chain (Koonin, 1991). The intrinsic disorder analysis revealed that RdRp protein possessed an ordered structure but contained virtually no IDPRs (Fig. 1H). Thus, the Hel protein has a well-predicted structure (Fig. 2H). Presence of several MoRFs was identified in the polypeptide chain by MoRFPred (Fig. 3). This indicates the essentiality of the RdRp domain (Agrawal et al., 2001). The identified high binding propensity of RdRp protein towards RNA residues suggests protein-RNA interactions (Fig. 4, 5, 6). This is consistent with the previous report which demonstrated the essentiality of the RdRp domain in viral replication (O' Reilly and Kao, 1998).

3.2 Analysis of ORF2 protein

The ORF2 encodes the capsid protein which forms the HEV virion major structural component (Chandra et al., 2008; Mori and Matsuura, 2011). The disorder profile of ORF2 protein showed significant disorder at the N-terminus with possession of multiple flexible regions (Fig. 2A). The protein-binding propensity analysis revealed the protein consisted of several MoRFs, which were distributed through its entire polypeptide chain length (Fig. 3). This indicated its vital role in protein function through mediating the interaction with the genomic RNA. Presence of many RNA-binding residues as predicted by DisoRDPbind and PPRInt in ORF2 protein at the N-terminus suggests its binding activity with the HEV RNA (Fig. 4, 6). This observation correlates with the fact that ORF2 (110 amino acids at N-terminus) interacts with the 5' region (encapsidation signal) of HEV genomic RNA (Surjit et al., 2004). Additionally, ORF2 N-terminus entails a signal peptide followed by an arginine-rich domain, which has shown involvement in viral RNA encapsidation during the assembly process (Surjit et al., 2004; Tam et al., 1991). Few DNA-binding residues at C-terminus by DisoRDPbind (Fig. 4) and several DNA-binding residues by DRNAPred (Fig. 5) were identified. It has been suggested that capsid formation is triggered through oligomerization of ORF2, when the specific region of ORF2 protein binds to the encapsidation signal. However, due to lack of cell culture models the underlying mechanisms remain to be completely understood.

3.3 Analysis of ORF3 protein

ORF3 is a phosphoprotein consisting of about 113 to 114 amino acid residues that performs a crucial function in the virus egress or release from infected cells (Emerson et al., 2006; Yamada et al., 2009).The graph depicting the disorder propensity in the ORF3 protein of HEV shows it to be a highly disordered protein (Fig. 2B). MoRFPred predicted the protein-binding residues at C-terminal end of the ORF3 protein (Fig. 3). Numerous RNA- and DNA-binding residues were identified in the ORF3 using DisoRDPbind and DRNAPred webservers, respectively (Fig. 4, 5). The ORF3 protein has been demonstrated to participate in interaction with several host proteins in addition to interaction with the ORF2 protein (Zafrullah et al., 1997). Additionally, the proline-rich motif (PSAP) in C-terminal region has been reported to play a role in the activity of ESCRT machinery by interacting with Tsg101 (Surjit et al., 2006).

Thus, our observations lead to the conclusions that ORF3 has the highest prevalence of disordered residues followed by ORF2, which has comparatively lesser fraction of intrinsic disorder. The ORF1 has the least number of disordered residues in the HEV proteome (Fig. S1).

4 Discussion

The discovery of the IDPs/IDPRs abundance in viral proteomes has emerged as a surprise to researchers. The preponderance of IDPRs in functionally active proteins (lacking properly folded structures) was considered as extraordinarily unique. This novel category of biologically active proteins was thus defined with the "prevalence of exclusivity/exceptionality" concept (Uversky, 2015). The advancement in computational tools has presented an opportunity for reliable prediction of intrinsic disorder predisposition in the proteomes, based on protein sequences. Thus, we have focused on disordered regions of the YDR by exploiting different bioinformatics algorithms for the prediction of intrinsic disorder predisposition. Further, Molecular recognition features (MoRFs) (protein-binding regions) were analysed in the HEV proteome. MoRFs are short disordered segments in IDPs/IDPRs that are prone to interactions with their binding partners upon transition from a disorder-to-order state (Metallo, 2010). Moreover, in addition to protein-binding related functions, intrinsically disordered segments also mediate functions by facilitating their interactions with nucleotides (DNA and RNA) (Peng et al., 2017). Thus, nucleotide-binding regions were also identified in the study.

The first domain (N-terminus) encoded byORF1 polyprotein is the Met domain. The disorder graph profile suggests Met as an ordered domain. The presence of MoRF in this particular domain protein reveals its interaction for the processing of ORF1 polyprotein. Few predicted nucleotide binding residues further suggests that Met domain assist in interactions. The second domain situated at N-terminus after methyltransferase is the Y domain by ORF1-encoded polyprotein. The intrinsic disorder analysis revealed it as a structured domain with multiple flexible regions. The essentiality of Y domain in cytoplasmic membrane binding has been established (Parvez, 2017). Thus, presence of MoRF region in this conserved counterpart clearly suggests the Y-domain essentiality in binding which support the previous literature (Parvez, 2017). Several RNA-binding residues at the C-terminus (including conserved counterpart) clearly indicate its essentiality in binding with the viral RNA. Moreover, identification of several DNA-binding residues showed high propensity for DNA to bind to Y domain protein. Thus, our results show consistency with the published reports showing the indispensability of the Y-domain in RNA replication (Parvez, 2017; Shafat et al., 2021). The PCP domain is positioned immediate downstream to the Y-domain. The disorder profile of PCP domain obtained shows it to be an IDPR. However, only single MoRF (N-terminal) in conjunction with few nucleotide-binding sites were identified in this domain. The X domain is situated between the Pro domain and Hel domain. The disorder profile shows it be an ordered domain with flexible N- terminus suggesting higher tendency of X domain towards intrinsic disorder. The identification of several RNA-binding residues N-terminus (potential Appr1″-pase active site at) supports the previous study, suggesting its significant role in the replication of HEV (Parvez, 2015). Moreover, the identified propensity of the 'NxxNxxHxxGGG' segment towards both RNA and DNA residues suggests catalytic/regulatory function performed by the non-structural X-domain (Parvez, 2015). The HVR domain is positioned between PCP and Pro domains. Presence of MoRF regions in the HVR indicates the crucial role played by it in the interactive ability with its binding partners. Prediction of several DNA-binding residues showed the propensity of HVR residues propensity in binding to DNA. The proline domain is situated between the HVR and X domains. The intrinsic disorder analysis revealed it as a complete disordered structure, suggesting it as an IDP which supports the earlier finding (Purdy et al., 2012). Prediction of MoRF regions shows consistency with the earlier observations which suggest disorder-to-order state transition in the Pro region (Purdy et al., 2012). Furthermore, absence of identified RNA- and DNA-binding residues in the proline region clearly reveals the non-essentiality of the Pro domain in viral replication which is coherent with earlier report (Pudupakam et al., 2009; Pudupakam et al., 2011). The RNA helicase domain is situated between the X and RdRp domains. It

has been suggested that the Hel domain interacts with Met domain in RNA capping functioning, which indicates protein-protein interaction which is consistent with our predicted MoRFPred results (Karpe and Lole, 2010). Additionally, the presence of nucleotide-binding residues in conserved regions: Walker A (GVPGSGKS; aa 975–982) and Walker B (DEAP; aa 1029–1032), substantiate our findings. The last domain encoded by ORF1 polyprotein is the RdRp domain. Presence of several MoRFs in the polypeptide chain indicates the essentiality of the RdRp domain which is consistent with the earlier report (Agrawal et al., 2001). The identified high binding propensity of RdRp protein towards RNA residues suggests protein-RNA interactions which is consistent with the previous report that demonstrated the essentiality of the RdRp domain in viral replication (O' Reilly 1998). Moreover, it has been suggested that RdRp binds to the 3' end of the genomic RNA and needs two stem–loop structures at the 3' end of the poly(A) stretch for this binding (Agrawal et al., 2001).

Presence of several MoRFs distributed through the entire polypeptide chain of ORF2 indicates its vital role in protein function through mediating the interaction with the genomic RNA. Presence of several RNA-binding residues suggests its binding activity with the HEV RNA. These findings are coherent with earlier investigations (Surjit e al., 2004; Tam et al., 1991). In ORF3 protein, MoRFs in conjunction with several nucleotide-binding residues were observed. The presence of MoRFs at C-terminal indicates crucial role played by ORF3 in the functioning of ORF3 protein (Surjit et al., 2006).

Thus, our intrinsic disorder analysis results revealed that ORF1 polyprotein consists of mostly ordered domains, i.e., proteins having significant level of well-defined structures, with the exception of Pro and PCP domains. The analysis reveals Pro domain as a highly disordered protein while PCP domain as an intrinsic disordered protein. MoRF analysis reveals that HEV proteome contains multiple MoRFs across all ORFs (at least one MoRF in proteins). Thus, these regions due to possession of MoRFs can be used for protein–protein interactions due to the structural flexibility. This clearly indicates significant role played by disorder in the functionality of the HEV structural and non-structural proteins. These results are indicative of molecular interactions (protein binding, RNA binding, and DNA binding) with the receptor on the host cell membrane and further viral infection. As IDPs/IDPRs are critical for molecular recognition, and therefore, can be considered as disorder-based targets in drug discovery.

5 Conclusions

Despite a major public health concern in the developing as well as developed countries, detailed information regarding the structural and non-structural protein structures and functions of HEV are precisely understood. Current study provides novel information on the prevalence of intrinsic disordered regions in the HEV proteome. These observations further strengthen the essentiality of IDPRs and noteworthy multifunction associated with it. This data on the prevalence of IDPRs in the HEV proteome could assist us in providing the pathways regarding the designing of novel drugs. However, the theoretical understanding based on the present data on IDPRs requires further experimental validation. Therefore, it is vital to reveal more information about the characteristics of IDPRs including its sequence, structure, dynamics, interactions network. Further our work will assist in elucidating the basic biology of HEV which is currently debated. More in-depth experimental studies using molecular and cell biology techniques to establish structure–function relationships are required for a better understanding of the functioning of HEV proteins.

Abbreviations

HEV Hepatitis E virus
IDP Intrinsically disordered proteins

IDPRs Intrinsically disordered protein regions
MoRFs Molecular recognition features
PONDR Predictor of natural disordered regions
PPRInt Prediction of protein RNA-interaction
Met Methyltransferase
PCP Putative papain-like cysteine protease
HVR Hypervariable region
Pro Proline region
Hel RNA helicase
RdRp RNA-dependent RNA polymerase

Supplementary Figures

Fig. S1 HEV ORF1 proteins with the predicted secondary structure: (A) Met domain; (B) Y-domain; (C) PCP domain; (D) HVR domain; (E) Pro domain; (F) X-domain; (G) Hel domain; and (H) RdRp domain. The analysis was conducted using Phyre2 webserver.

Fig. S2 HEV ORF2 protein with the predicted secondary structure. The analysis was conducted using Phyre2 webserver.

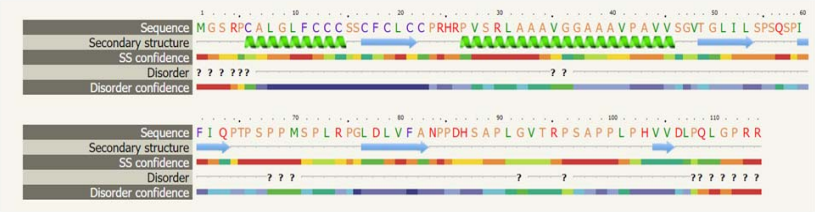

Fig. S3 HEV ORF3 protein with the predicted secondary structure. The analysis was conducted using Phyre2 webserver.

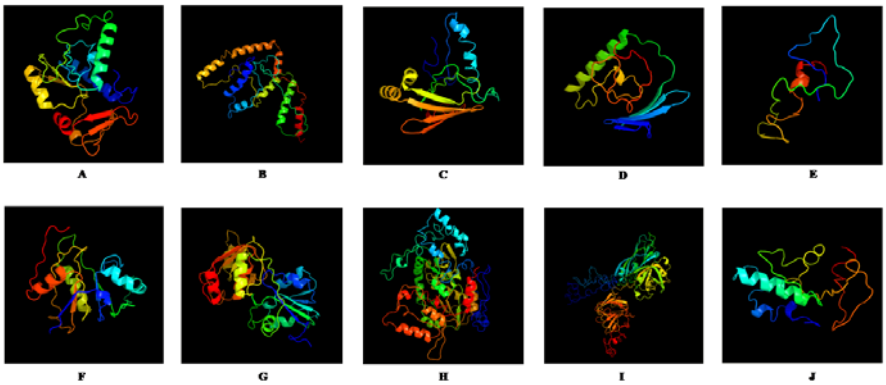

Fig. S4 HEV proteins with the predicted tertiary structure: (A) Met domain; (B) Y-domain; (C) PCP domain; (D) HVR domain; (E) Pro domain; (F) X-domain; (G) Hel domain; and (H) RdRp domain; (I) ORF2; and (J) ORF3. The analysis was conducted using Phyre2 webserver.

Fig. S5 Analysis of protein-binding propensity of HEV proteome. The resulting protein-binding profile was calculated using DisoRDPbind webserver. The identified RNA-binding residues are depicted in red.

Acknowledgment

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

Agrawal S, Gupta D, Panda SK. 2001. The 30 end of hepatitis E virus (HEV) genome binds specifically to the viral RNA-dependent RNA polymerase (RdRp). Virology, 282: 87-101

Anang S, Subramani C, Nair VP, Kaul S, Kaushik N, Sharma C, Tiwari A, Ranjith-Kumar CT, Surjit M. 2016.

Identification of critical residues in Hepatitis E virus macro domain involved in its interaction with viral methyltransferase and ORF3 proteins. Scientific Reports, 6(1): 1-3

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

Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2016. GenBank. Nucleic Acids Research, 45(D1): D37-D42

Borkakoti J, Ahmed G, Rai A, Kar P. 2017. Report of novel H105R, D29N, V27A mutations in the methyltransferase region of the HEV genome in patients with acute liver failure. Journal of Clinical Virology, 91: 1-4

Chandra V, Taneja S, Kalia M, Jameel S. 2008. Molecular biology and pathogenesis of hepatitis E virus. Journal of Biosciences, 33(4): 451-464

Devhare P, Sharma K, Mhaindarkar V, Arankalle V, Lole K. 2014. Analysis of helicase domain mutations in the hepatitis E virus derived from patients with fulminant hepatic failure: effects on enzymatic activities and virus replication. Virus Research, 184: 103-110

Ding Q, Heller B, Capuccino JM, Song B, Nimgaonkar I, Hrebikova G, Contreras JE, Ploss A. 2017. Hepatitis E virus ORF3 is a functional ion channel required for release of infectious particles. Proceedings of the National Academy of Sciences of the United States of America, 114(5): 1147-1152

Disfani FM, Hsu WL, Mizianty MJ, Oldfield CJ, Xue B, Dunker AK, Uversky VN, Kurgan L. 2012. MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. Bioinformatics, 28: i75-83

Dosztanyi Z, Chen J, Dunker AK, Simon I, Tompa, P. 2006. Disorder and sequence repeats in hub proteins and their implications for network evolution. Journal of Proteome Research, 5: 2985-2995

Dunker AK, Brown CJ, Obradovic Z. 2002. Identifcation and functions of usefully disordered proteins. Advances in Protein Chemistry, 62: 25-49

Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks. The FEBS Journal, 272: 5129-5148

Dunker AK, Garner E, Guilliot S, Romero P, Albrecht K, Hart J, Obradovic Z, Kissinger C, Villafranca JE. 1998. Protein disorder and the evolution of molecular recognition: Theory, predictions and observations. Pacific Symposium on Biocomputing, 473-484

Dunker AK, Oldfield CJ, Meng J, Romero P, Yang JY, Chen JW, Vacic V, Obradovic Z, Uversky VN. 2008. The unfoldomics decade: an update on intrinsically disordered proteins. BMC Genomics, 9(2): 1-26

Dunker AK, Silman I, Uversky VN, Sussman JL. 2008. Function and structure of inherently disordered proteins. Current Opinion in Structural Biology, 18: 756-764

Emerson SU, Nguyen H, Torian U, Purcell RH. 2006. ORF3 protein of hepatitis E virus is not required for replication, virion assembly, or infection of hepatoma cells in vitro. Journal of Virology, 80: 10457-10464

Giri R, Kumar D, Sharma N, Uversky VN. 2016. Intrinsically disordered side of the Zika virus proteome. Frontiers in Cellular and Infection Microbiology, 6: 144

He M, Wang M, Huang Y, Peng W, Zheng Z, Xia N, Xu J, Tian D. 2016. The ORF3 protein of genotype 1 hepatitis E virus suppresses TLR3-induced NF-κB signaling via TRADD and RIP1. Scientific Reports, 6(1): 1-3

Kabrane-Lazizi Y, Meng XJ, Purcell RH, Emerson SU. 1999. Evidence that the genomic RNA of hepatitis E virus is capped. Journal of Virology, 73(10): 8848-8850

Kadare G, Haenni AL. 1997. Virus-encoded RNA helicases. Journal of Virology, 71: 2583-2590

Kanade GD, Pingale KD, Karpe YA. 2018. Activities of Thrombin and Factor Xa Are Essential for Replication of Hepatitis E Virus and Are Possibly Implicated in ORF1 Polyprotein Processing. Journal of Virology, 92: e01853-17

Karpe YA, Lole KS. 2010. NTPase and 5′ to 3′ RNA duplex-unwinding activities of the hepatitis E virus helicase domain. Journal of Virology, 84(7): 3595-3602

Karpe YA, Lole KS. 2010. RNA 5′-triphosphatase activity of the hepatitis E virus helicase domain. Journal of Virology, 84(18): 9637-9641

Koonin EV, Gorbalenya AE, Purdy MA, Rozanov MN, Reyes GR, Bradley DW. 1992. Computer-assisted assignment of functional domains in the nonstructural polyprotein of hepatitis E virus: delineation of an additional group of positive-strand RNA plant and animal viruses. Proceedings of the National Academy of Sciences of the United States of America, 89(17): 8259-8263

Koonin EV. 1991. The phylogeny of RNA-dependent RNA polymerases of positive strand RNA viruses. Journal of General Virology, 72(9): 2197-2206

Kumar M, Gromiha MM, Raghava GPS. 2008. Prediction of RNA binding sites in a protein using SVM and PSSM profle. Proteins, 71: 189-194

LeDesma R, Nimgaonkar I, Ploss A. 2019. Hepatitis E virus replication. Viruses, 11(8): 719

Liu J, Perumal NB, Oldfeld CJ, Su EW, Uversky VN, Dunker AK. 2006. Intrinsic disorder in transcription factors. Biochemistry, 45: 6873-6888

Magden J, Takeda N, Li T, Auvinen P, Ahola T, Miyamura T, Merits A, Kääriäinen L. 2001. Virus-specific mRNA capping enzyme encoded by hepatitis E virus. Journal of Virology, 75(14): 6249-6355

Mahilkar S, Paingankar MS, Lole KS. 2016. Hepatitis E virus RNA-dependent RNA polymerase: RNA template specificities, recruitment and synthesis. Journal of General Virology, 97(9): 2231-2242

Metallo SJ. 2010. Intrinsically disordered proteins are potential drug targets. Current Opinion in Chemical Biology, 14(4): 481-488

Midic U, Oldfield CJ, Dunker AK, Obradovic Z, Uversky VN. 2009. Protein disorder in the human diseasome: unfoldomics of human genetic diseases. BMC Genomics, 10(1): 1-24

Mori Y, Matsuura Y. 2011. Structure of hepatitis E viral particle. Virus Research, 161: 59-64

Nan Y, Zhang YJ. 2016. Molecular biology and infection of hepatitis E virus. Frontiers in Microbiology, 7: 1419

O'Reilly EK, Kao CC. 1998. Analysis of RNA-dependent RNA polymerase structure and function as guided by known polymerase structures and computer predictions of secondary structure. Virology, 252: 287-303

Ojha NK, Lole KS. 2016. Hepatitis E virus ORF1 encoded macro domain protein interacts with light chain subunit of human ferritin and inhibits its secretion. Molecular and Cellular Biochemistry, 417(1-2): 75-85

Oldfeld CJ, Dunker AK. 2014. Intrinsically disordered proteins and intrinsically disordered protein regions. Annual Review of Biochemistry, 83: 553-584

Paliwal D, Panda SK, Kapur N, Varma SP, Durgapal H. 2014. Hepatitis E virus (HEV) protease: A chymotrypsin-like enzyme that processes both non-structural (pORF1) and capsid (pORF2) protein. Journal of General Virology, 95: 1689-1700

Parvez MK, Al-Dosari MS. 2015. Evidence of MAPK-JNK1/2 activation by hepatitis E virus ORF3 protein in cultured hepatoma cells. Cytotechnolog, 67(3): 545-550

Parvez MK, Khan AA. 2014. Molecular modeling and analysis of hepatitis E virus (HEV) papain-like cysteine protease. Virus Research, 179: 220-224

Parvez MK. 2013. Molecular characterization of hepatitis E virus ORF1 gene supports apapain-like cysteine protease (PCP)-domain activity. Virus Research, 178(2): 553-556

Parvez MK. 2015. The hepatitis E virus ORF1 'X-domain' residues form a putative macrodomain protein/Appr-1"-pase catalytic-site, critical for RNA replication. Gene, 566(1): 47-53

Parvez MK. 2017. Mutational analysis of hepatitis E virus ORF1 'Y-domain': effects on RNA replication and virion infectivity. World Journal of Gastroenterology, 23(4): 590-602

Parvez MK. 2017. Mutational analysis of hepatitis E virus ORF1 'Y-domain': effects on RNA replication and virion infectivity. World Journal of Gastroenterology, 23(4): 590-602

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

Pehrson JR, Fried VA. 1992. MacroH2A, a core histone containing a large nonhistone region. Science, 257:1398–1400

Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK, Obradovic Z. 2005. Optimizing long intrinsic disorder predictors with protein evolutionary information. J Bioinformatics and Computational Biology, 3: 35-60

Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK, Obradovic Z.2006. Length-dependent prediction of protein intrinsic disorder. BMC Bioinformatics, 7: 208

Peng Z, Wang C, Uversky VN, Kurgan L (2017) Prediction of disordered RNA, DNA, and protein binding regions using DisoRDPbind. Methods in Molecular Biology, 1484: 187-203

Peng ZKL. 2015. High-throughput prediction of RNA, DNA and protein binding regions mediated by intrinsic disorder. Nucleic Acids Research, 43: e121

Proudfoot A, Hyrina A, Holdorf M, Frank AO, Bussiere D. 2019. First crystal structure of a nonstructural hepatitis E viral protein identifies a putative novel zinc-binding protein. Journal of Virology, 93(13)

Pudupakam RS, Huang YW, Opriessnig T, Halbur PG, Pierson FW, Meng XJ. 2009. Deletions of the hypervariable region (HVR) in open reading frame 1 of hepatitis E virus do not abolish virus infectivity: evidence for attenuation of HVR deletion mutants in vivo. Journal of Virology, 83(1): 384-395

Pudupakam RS, Kenney SP, Cordoba L, Huang YW, Dryman BA, LeRoith T, Pierson FW, Meng XJ. 2011. Mutational analysis of the hypervariable region of hepatitis E virus reveals its involvement in the efficiency of viral RNA replication. Journal of Virology, 85(19): 10031-10040

Purdy MA, Lara, J, Khudyakov YE. 2012. The hepatitis E virus polyproline region is involved in viral adaptation. PLoS ONE 7: e35974

Purdy MA. 2012. Evolution of the hepatitis E virus polyproline region: order from disorder. Journal of Virology, 86: 10186-10193

Salma P, Chhatbar C, Seshadri S. 2009. Intrinsically unstructured proteins: Potential targets for drug discovery. American Journal of Infectious Diseases, 5: 133-141

Sehgal D, Thomas S, Chakraborty M, Jameel S. 2006. Expression and processing of the Hepatitis E virus ORF1 nonstructural polyprotein. Virology Journal, 3: 38

Shafat Z, Hamza A, Islam A, Al-Dosari MS, Parvez MK, Parveen S. 2021. Structural exploration of Y-domain reveals its essentiality in HEV pathogenesis. Protein Expression and Purification, 187: 105947

Singh A, Kumar A, Yadav R, Uversky VN, Giri R. 2018. Deciphering the dark proteome of Chikungunya virus. Scientific Reports, 8: 5822

Smith DB, Simmonds P. 2015 Hepatitis E virus and fulminant hepatitis- a virus or hostspecific pathology? Liver International, 35(4): 1334-1340

Surjit M, Jameel, S, Lal SK. 2004. The ORF2 protein of hepatitis E virus binds the 50 region of viral RNA. Journal of Virology, 78: 320-328

Surjit M, Oberoi R, Kumar R, Lal SK. 2006. Enhanced alpha(1) microglobulin secretion from hepatitis E virus ORF3-expressing human hepatoma cells is mediated by the tumor susceptibility gene 101. Journal of

Biological Chemistry, 281: 8135-8142

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

Tam AW, Smith MM, Guerra ME. 1991. Hepatitis E virus (HEV): Molecular cloning and sequencing of the full-length viral genome. Virology, 185(1): 120-131

Tsarev SA, Emerson SU, Reyes GR, Tsareva TS, Legters LJ, Malik IA, Iqbal M, Purcell RH (1992) Characterization of a prototype strain of hepatitis E virus. Proceedings of the National Academy of Sciences of the United States of America, 89: 559-563

Uversky VN (2015) Paradoxes and wonders of intrinsic disorder: prevalence of exceptionality. Intrinsically Disordered Proteins, 3: e1065029

Uversky VN, Oldfield CJ, Dunker AK. 2005. Showing your ID: Intrinsic disorder as an ID for recognition, regulation and cell signaling. Journal of Molecular Recognition: An Interdisciplinary Journal, 18: 343-384

Van Der Lee R, Buljan M, Lang B et al. 2014. Classifcation of intrinsically disordered regions and proteins. Chemical Reviews, 114: 6589-6631

Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. 2004. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. Journal of Molecular Biology, 337: 635-645

Wright PE, Dyson HJ. 1999. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. Journal of Molecular Biology, 293:321–331

Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN. 2010. PONDR-Fit: A meta-predictor of intrinsically disordered amino acids. Biochimica et Biophysica Acta, 1804(4): 996-1010

Xue B, Williams RW, Oldfield CJ, Goh GK, Dunker AK, Uversky VN. 2010. Viral disorder or disordered viruses: do viral proteins possess unique features? Protein and Peptide Letters, 17: 932-951

Yamada K, Takahashi M, Hoshino Y, Takahashi H, Ichiyama K, Nagashima S, Tanaka T, Okamoto H. 2009. ORF3 protein of hepatitis E virus is essential for virion release from infected cells. Journal of General Virology, 90: 1880-1891

Yan J, Kurgan L, 2017. DRNApred, fast sequence-based method that accurately predicts and discriminates DNA- and RNA-binding residues. Nucleic Acids Research, e84-e84

Zafrullah M, Ozdener MH, Panda SK, Jameel S. 1997. The ORF3 protein of hepatitis E virus is a phosphoprotein that associates with the cytoskeleton. Journal of Virology, 71: 9045-9053

Zhang M, Purcell RH, Emerson SU. 2001. Identification of the 5′ terminal sequence of the SAR-55 and MEX-14 strains of hepatitis E virus and confirmation that the genome is capped. Journal of Medical Virology, 65(2): 293-295