Dark Proteome of Newly Emerged SARS-CoV-2 in Comparison with Human and Bat Coronaviruses

Rajanish Giri[1]*, Taniya Bhardwaj[1], Meenakshi Shegane[1], Bhuvaneshwari R. Gehi[1], Prateek Kumar[1], Kundlik Gadhave[1]

[1]Indian Institute of Technology Mandi, School of Basic Sciences, VPO Kamand, Himachal Pradesh 175005, India

***Correspondence to:** Dr. Rajanish Giri, School of Basic Sciences, Indian Institute of Technology Mandi, Himachal Pradesh 175005, India. Email: rajanishgiri@iitmandi.ac.in. Telephone number: 01905-267134, Fax number: 01905-267138

Abstract:

Recently emerged Wuhan's novel coronavirus designated as SARS-CoV-2, a causative agent of coronavirus disease 2019 (COVID-19) is rapidly spreading its pathogenicity throughout the world now. More than 4000 mortalities have occurred worldwide till the writing of this article and this number is increasing every passing hour. World Health Organization (WHO) has declared it as a global public health emergency. The multiple sequence alignment data correlated with already published reports on SARS-CoV-2 indicated that it is closely related to Bat-Severe Acute Respiratory Syndrome like coronavirus (Bat CoV SARS-like) and well-studied Human SARS. In this study, we have exploited the complementary approach to examine the intrinsically disordered regions in proteome of SARS-CoV-2 using Bat SARS-like and Human SARS CoVs as comparative models. According to our findings, SARS-CoV-2 proteome contains a significant amount of ordered proteins except Nucleocapsid, Nsp8, and ORF6. Further, cleavage sites in replicase 1ab polyprotein are found to be highly disordered. We have extensively investigated the dark proteome in SARS-CoV-2 which will have implications for the structured and unstructured biology of SARS or SARS-like coronaviruses.

Significance: The infection caused by Novel Coronavirus (SARS-CoV-2) is responsible for the current pandemic that cause severe respiratory disease and pneumonia-like infection in humans. Currently, there is no such in-depth information on protein structure and function available in public domain and no effective anti-viral or vaccines are available for the treatment of this infection. Our study provides comparative order and disorder-based proteome information with Human SARS and Bat CoV that may be useful for structure-based drug discovery.

Keywords: COVID-19, Coronavirus, Bat CoV, Human SARS, Intrinsically Disordered Proteins

List of Abbreviations:

ACE2	Angiotensin-converting enzyme 2

CDF	Cumulative distribution function
CH	Charge hydropathy
COVID-19	Coronavirus disease 2019
CTD	C-terminal domain
DMVs	Double-membrane vesicles
ICTV	International committee on taxonomy of viruses
IDP	Intrinsically disordered proteins
IDPRs	Intrinsically disordered protein regions
IFN	Interferon
MSA	Multiple sequence alignment
Nsps	Non-structural proteins
NTD	N-terminal domain
PONDR	Predictor of natural disordered regions
PPID	Predicted percentage of intrinsic disorder
RBD	Receptor binding domain
SARS	Severe acute respiratory syndrome
TRS	Transcriptional regulatory sequences
VLPs	Virus-like particles
WHO	World health organization

Introduction:

Emerging Coronavirus disease 2019 (COVID-19) is a recent pandemic and declared as a public health emergency by World Health Organization (WHO). Since its first case in December 2019 in Wuhan's seafood and meat market, China, it has now large scale socio-economic impact globally (Yang *et al*, 2020). According to WHO, till 10th March 2020, the infections reached 109 countries (more than 0.1 million confirmed cases). Its symptoms include gastrointestinal tract complications, upper respiratory problems such as cough, cold, shortness of breath, and fever, and in some lethal conditions can cause mortality (Coronavirus disease 2019).

According to the norms of the International Committee on Taxonomy of Viruses (ICTV), SARS-CoV-2 comes under coronavirinae sub-family of *coronaviridae* family of order nidovirales. Nidovirales order viruses are enveloped, non-segmented positive-sense, single-stranded RNA viruses (Gorbalenya *et al*, 2006). The major variations among *Nidovirus* family occur in the number, type, and sizes of the structural proteins (Gorbalenya *et al*, 2006). The family *coronaviridae* comprises of vertebrate infecting viruses that transmit horizontally mainly through oral/fecal route and cause gastrointestinal and respiratory problems to the host (Corman *et al*, 2019). Sub-family *coronavirinae* consists of four genera, namely: alpha, beta, gamma, and delta coronaviruses based on the phylogenetic clustering of viruses (Woo *et al*, 2012; Cotten *et al*, 2013). *Coronavirinae* having the largest genomes for RNA viruses, incorporate ~30kb genomes inside the enveloped capsid. These viruses have variations among genome due to significant alterations in the structure and morphology of the nucleocapsids of virions (Masters, 2006).

SARS-Coronavirus genomic RNA accommodates a 5' cap, leader sequence, UTR, a replicase gene, genes for structural and accessory protein, 3' UTR, and a poly-A tail (**Figure 1**). Two-

third of genome codes for replicase gene (~20kb) containing genes for all non-structural viral proteins while the rest ~10kb genome contains genes for accessory proteins interspersed between the genes responsible for coding structural proteins (Masters, 2006). For the expression of structural and accessory proteins, additional transcriptional regulatory sequences (TRS) are present within the viral genome. 5' and 3' UTRs contain stem-loop regions required for viral RNA synthesis (Hussain *et al*, 2005). The ~20kb (replicase gene) ssRNA is translated first into two long polyproteins: replicase polyprotein 1a and 1ab inside host cells. The newly formed polyproteins after cleavage by two viral proteases result in 16 non-structural proteins that perform a wide range of functions for viruses inside the host cell. They also induce ER-derived double-membrane vesicles (DMVs) for viral replication and transcription. Structural proteins shape the outer cover of the virion, while accessory proteins are mostly involved in host immune evasion (Snijder *et al*, 2006; Sawicki *et al*, 2007).

In this manuscript, we want to analyze the dark proteome of the SARS-CoV-2, so explaining a perspective on ordered and disordered proteome is essential. In classical structure-function-paradigm, a stable and defined 3-Dimensional structure is a prerequisite for a protein to accomplish its role. This notion dominated over the years before the idea of intrinsic disorder in protein structures came into full discussions and acceptance among the structural biologists. The other class of proteins fails to fold into well-defined structures and instead remain disordered under physiological buffer conditions. These proteins are called intrinsically disordered proteins (IDPs). Regions within proteins that contain unstructured segments are called intrinsically disordered protein regions (IDPRs). The property of being intrinsically disordered in proteins is determined by the amino acid sequences (Van Der Lee *et al*, 2014; Oldfield & Dunker, 2014; Wright & Dyson, 1999). IDPs exhibit their biological functions in signaling, gene regulation, and control processes by interacting with their physiological partners (Dunker *et al*, 2005, 2002, 2008; Liu *et al*, 2006; Uversky *et al*, 2005).

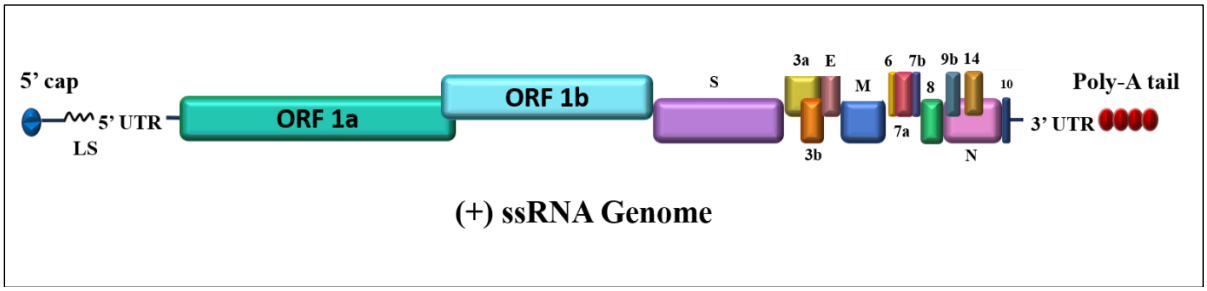

Figure 1. Genome architecture of SARS-CoV-2. SARS-CoV-2 contains a positive single-stranded mRNA as genetic material. The 5' capped mRNA has a leader sequence (LS), poly-A tail at 3' end, and 5' and 3' UTR. It contains following genes: ORF1a, ORF1b, Spike (S), ORF3a, ORF3b, Envelope (E), Membrane (M), ORF6, ORF7a, ORF7b, ORF8, ORF9b, ORF14, Nucleocapsid (N), and ORF10.

Zhang and colleagues have reported the genomic sequence of the novel Wuhan seafood pneumonia virus or SARS-CoV-2 with GenBank accession number NC_045512 having 29,903 nucleotides. It was isolated from the bronchoalveolar lavage fluid of a patient and has been named finally as SARS-CoV-2 by WHO (Wu *et al*, 2020b).

As the IDPs are present in all three kingdoms of life, viral proteins often contain unstructured regions that have been strongly correlated with their virulence (Xue *et al*, 2010b; Giri *et al*, 2016; Singh *et al*, 2018; Ward *et al*, 2004). In this report, we designed our study to investigate the disordered side of the novel SARS-CoV-2 using computational approaches. We also have comprehensively analyzed the disordered regions among the closely related viruses; Human SARS and Bat CoV (SARS-like) like coronaviruses. As all three viruses have high similarity, therefore, it is a great approach to understand the structural and sequence properties. We believe that this study will help the structural and non-structural biologists to perform experiments for a more in-depth understanding of this virus. This will have certainly long-term implications for developing the new drugs or vaccines against this currently unpreventable infection.

Materials and Methods:

Sequence retrieval and multiple sequence alignment: The protein sequences of Bat CoV (SARS-like), and Human SARS CoV were retrieved from UniProt (UniProt IDs for individual proteins are listed in the **table 1**). The translated sequences of SARS-CoV-2 proteins [GenBank database (Clark *et al*, 2016) (Accession ID: NC_045512.2)] were obtained from GenBank. We used these sequences for performing multiple sequence alignment (MSA) and predicting the IDPRs. We have used Clustal Omega (Sievers *et al*, 2011) for protein sequence alignment and Esprit 3.0 (Robert & Gouet, 2014) for constructing the aligned images.

Prediction of disordered regions: For prediction of intrinsic disorder in SARS virus proteomes, we have used multiple predictors: PONDR® (Predictor of Natural Disordered Regions) consortium [PONDR®VLS2 (Peng *et al*, 2006), PONDR®VL3 (Peng *et al*, 2005), PONDR®FIT (Xue *et al*, 2010a), PONDR® VLXT (Romero *et al*, 2001), IUPRED [Long and Short] (Mészáros *et al*, 2018). These artificial neural networks predict the residues which do not possess the tendency to form a structure in isolation. Additionally, two constructs of the IUPRED tool have also been utilized for the prediction of long and short IDPRs in query proteins. Residues with disorder score more than the 0.5 threshold values are considered as intrinsically disordered residues, whereas residues with the predicted disorder scores between 0.2 and 0.5 are considered flexible. Complete predicted percent of intrinsic disorder (PPID) was calculated for every protein of all the three viruses from outputs of six predictors. The detailed methodology has been given in our previous reports (Gadhave *et al*, 2020; Garg *et al*, 2019).

Combined CH-CDF analysis to predict disorder predisposition of proteins: Charge hydropathy plot (Uversky *et al*, 2000) and cumulative distribution function are two binary predictors of disorder which are available on the PONDR web server (https://www.pondr.com/). Combining the result from these binary predictors helps analyze the intrinsic disorder propensity of a protein-based on its charge, hydropathy and PONDR®VLXT score (Huang *et al*, 2012).

Results and Discussion:

Comprehensive analysis of Intrinsic disorder in proteins of SARS-CoV–2, Human SARS and Bat CoV (SARS-like):

The Mean predicted percentage of intrinsic disorder (Mean PPID) scores, that were obtained by averaging the predicted disorder scores from six disorder predictors (**Supplementary Table 1-3**) for each protein of SARS-CoV-2 as well as Human SARS, and Bat CoV have been represented in **table 1**.

Table 1: Evaluation of mean predicted percentage of intrinsic disorder (Mean PPID) in structural and accessory proteins of SARS-CoV-2, Human SARS, and Bat CoVs.

Proteins	**SARS-CoV-2**		**Human SARS CoV**		**Bat CoV (SARS-like)**	
	Length of protein (NCBI RefSeq accession ID)	**Mean PPID**	**Length of protein (UniProt ID)**	**Mean PPID**	**Length of protein (UniProt ID)**	**Mean PPID**
Spike glycoprotein (S)	1273 (YP_009724390.1)	1.41	1255 (P59594)	1.12	1242 (Q3LZX1)	1.85
Envelope (E)	75 (YP_009724392.1)	5.33	76 (P59637)	6.58	76 (Q3LZW9)	6.58
Membrane (M)	222 (YP_009724393.1)	2.70	221 (P59596)	1.36	221 (Q3LZX9)	1.36
Nucleocapsid (N)	419 (YP_009724397.2)	60.38	422 (P59595)	71.09	421 (Q3LZX4)	65.80
ORF3a	275 (YP_009724391.1)	9.09	274 (P59632)	8.76	274 (Q3LZX0)	6.20
ORF3b	22*	0	154 (P59633	7.14	39*	23.08
ORF6	61 (YP_009724394.1)	22.95	63 (P59634)	20.63	63 (Q3LZX8)	20.63
ORF7a	121 (YP_009724395.1)	1.65	122 (P59635)	0.82	122 (Q3LZX7)	0.82
ORF7b	43 (YP_009725296.1)	9.30	44 (Q7TFA1)	4.55	44 (Q3LZX6)	4.55

ORF8	121 (YP_009724396.1)	0.00	ORF8a – 39 (Q7TFA0)	2.56	121 (Q3LZX5)	0.00
			ORF8b – 84 (Q80H93	2.38		
ORF9b	97*	10.31	98 (P59636)	26.53	97 (Q3LZX3)	9.28
ORF10#	38 (YP_009725255.1)	0.00	-	-	-	-
ORF 14	73*	0	70 (Q7TLC7)	2.86	70*	0

#No sequence similarity was found on performing BLAST.

*These sequences are based on genome annotations done by Wu *et al.*, (Wu *et al*, 2020a).

Figures 2A, 2B and **2C** are 2D-disorder plots of SARS-CoV-2, Human SARS and Bat CoV respectively and represent the $PPID_{PONDR\text{-}FIT}$ vs $PPID_{Mean}$ plot. Based on the predicted levels of intrinsic disorder, proteins can be classified as highly ordered (PPID < 10%), moderately disordered (10% ≤ PPID < 30%) and highly disordered (PPID ≥ 30%) (Rajagopalan *et al*, 2011). From the data in **table 1, figures 2A, 2B** and **2C**; as well as the PPID based classification, we conclude that the Nucleocapsid protein from all three strains of coronavirus possesses the highest percentage of the disorder. ORF3b protein in Bat Cov, ORF6 protein in SARS-CoV-2, Human SARS, and Bat CoV and ORF9b protein in Human SARS and SARS CoV belong to the class of moderately disordered proteins. While the structured proteins, namely, Spike glycoprotein (S), Envelope protein (E) and Membrane protein (M) as well as accessory proteins ORF3a, ORF7a, ORF8 (ORF8a and ORF8b in case of Human SARS) of all three strains of coronavirus are ordered. ORF14 and ORF10 proteins also belong to the class of ordered proteins.

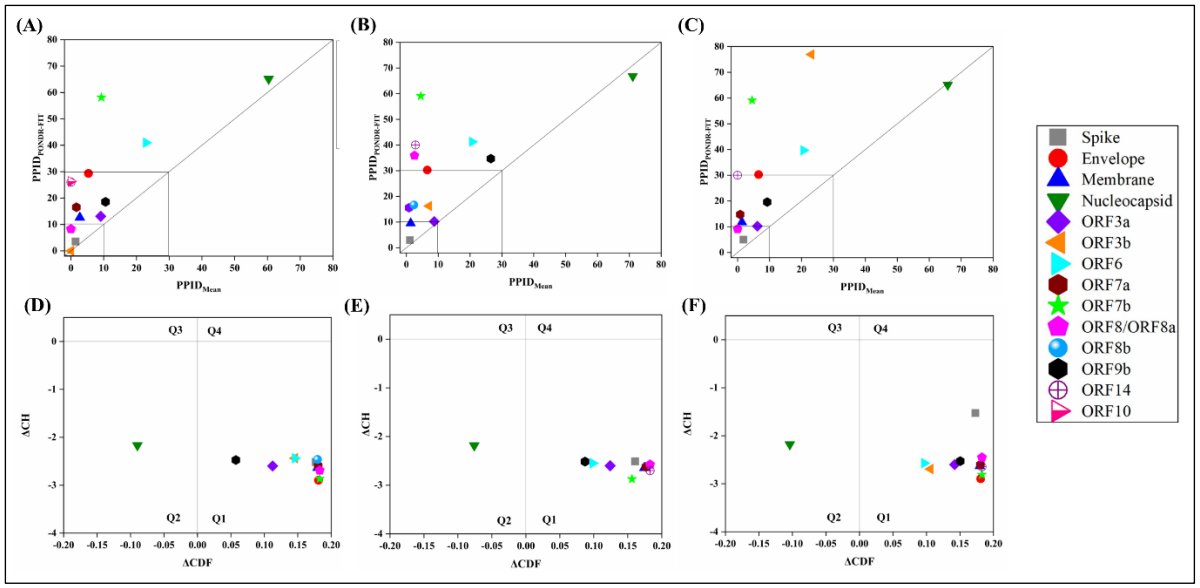

Figure 2. **Analysis of overall disorder status of proteins of SARS-CoV–2, Human SARS and Bat CoV (SARS-like)**: 2D plot representing $PPID_{PONDR\text{-}FIT}$ vs $PPID_{Mean}$ in **(A)** SARS-CoV-2 **(B)** Human SARS and **(C)** Bat CoV. In CH-CDF plot of the proteins of **(D)** SARS-CoV-2 **(E)** Human SARS and **(F)** Bat CoV, the Y coordinate of each protein spot signifies distance of corresponding protein from the boundary in CH plot and the X coordinate value corresponds to the average distance of the CDF curve for the respective protein from the CDF boundary.

In order to further investigate the nature of the disorder in proteins of SARS-CoV-2, Human SARS, and Bat CoV; the results obtained from two binary classifiers of disorder; Charge hydropathy (CH) plot and Cumulative distribution function (CDF) plot were combined. This helped evaluate the disorder predisposition of the proteins based on their charge, hydropathy and PONDR®VLXT disorder scores. The CH plot is a linear classifier that differentiates between proteins that are predisposed to possess extended disordered regions and include random coils and pre-molten globules from proteins that have compact conformations (ordered proteins and molten globule like proteins). The other binary predictor, CDF is a nonlinear classifier as it uses PONDR®VLXT scores to discriminate ordered globular proteins from all disordered conformations, which include, native molten globules, pre-molten globules, and random coils. The CH-CDF plot can be divided into four quadrants: Q1, which is expected to include ordered proteins; Q2, which will include proteins predicted to be disordered by CDF and ordered by CH; Q3, which represents proteins that are predicted to be disordered by both CH and CDF analysis and Q4, which includes proteins disordered according to CH but ordered according to CDF analysis (Gadhave *et al*, 2020). **Figures 2D, 2E** and **2F** represent the CH-CDF analysis of proteins of SARS-CoV-2, Human SARS, and Bat CoV and it can be observed that all the proteins are located within the two quadrants Q1 and Q2. The CH-CDF analysis leads to the conclusion that all proteins of SARS-CoV-2, Human SARS, and Bat CoV are ordered except Nucleocapsid protein, which is predicted to be disordered by CDF but ordered by CH and hence lies in Q2.

Intrinsic disorder analysis of structural proteins of Coronaviruses:

Coronaviruses encode four structural proteins, namely; Spike (S), Envelope (E) glycoproteins, Membrane (M), and Nucleocapsid (N), respectively. These proteins are translated from the last ~10kb nucleotides and form the outer cover of CoVs encapsulating their single-stranded genomic RNA.

Spike (S) glycoproteins: S protein is a multifunctional protein forming the exterior of SARS coronavirus particles (Cavanagh & Davis, 1986; Graham & Baric, 2010). It has two non-covalently joined subunits S1 and S2 which form surface homotrimers. Subunit S1 binds to host cell receptors, and, S2 mediates the viral entry into host cells (Belouzard *et al*, 2012; de Haan *et al*, 2006). Spike binds to the virion M protein through its C-terminal transmembrane region (McBride *et al*, 2007). Belonging to a class I viral fusion protein, S protein binds to specific surface receptor angiotensin-converting enzyme 2 (ACE2) on host cell plasma membrane through its N-terminus receptor-binding domain (RBD) and mediates viral entry into host cells (Li *et al*, 2005).

The protein consists of an N-terminal signal peptide, an extracellular domain, a transmembrane domain, and an intracellular domain (Broer *et al*, 2006). A 3.60 Å resolution structure (PDB ID: 6ACC) of S protein from Human SARS complexed with its host binding partner ACE2 has been obtained by cryo-electron microscopy (cryo-EM) (**Figure 3D**). In this PDB structure, few residues (1-17, 240-243, 661-673, 812-831 and 1120-1203) are missing (Song *et al*, 2018). Structure of S protein (3.5 Å) of SARS-CoV-2 has been recently deduced by Wrapp *et. al.*, using electron microscopy. The biophysical analysis in their study has also revealed that the protein has a higher binding affinity to ACE2 than S protein of Human SARS (Wrapp *et al*, 2020).

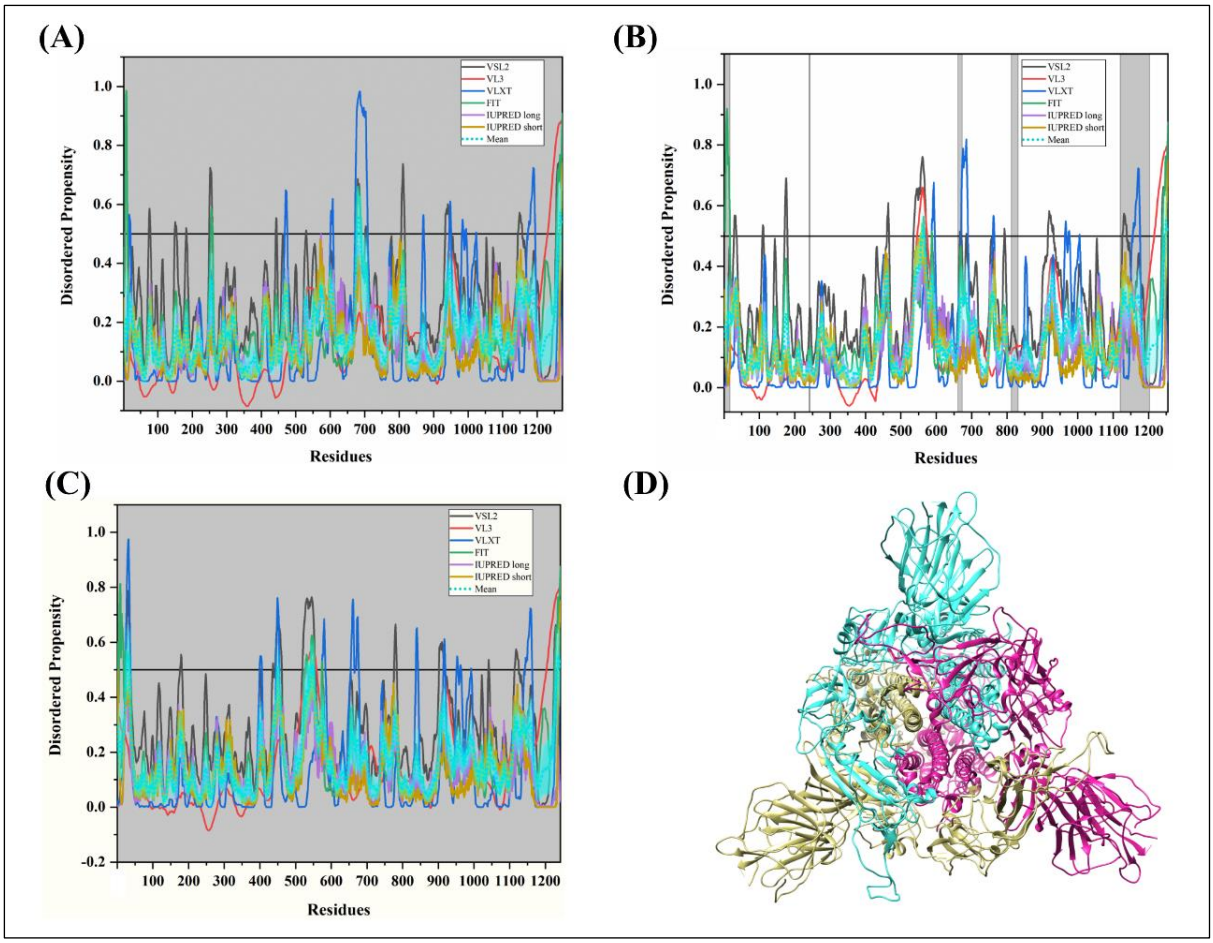

Figure 3. Intrinsic disorder propensity of Spike glycoprotein. (A) The graph represents the predicted IDP analysis in SARS-CoV-2, **(B)** Human SARS and **(C)** Bat CoV. Graphs in **(A)**, **(B)** and **(C)** show the disorder profiles generated by PONDR® VSL2 (black line), PONDR® VL3 (red line), PONDR® VLXT (blue line), PONDR® FIT (green line), IUPRED long (purple line) and IUPRED short (golden line). The mean disorder propensity which has been calculated by averaging the disorder scores from all six predictors is represented by short-dot line (sky-blue line) in the graph. The light sky-blue shadow region signifies the mean error distribution. The residues missing in the PDB structure or the residues for which PDB structure is unavailable is represented by the grey-coloured area in the graph. **(D)** A 3.60Å resolution PDB structure of spike glycoprotein of Human SARS obtained by cryo-EM (PDB ID: 6ACC) of 18-239, 244-660, 674-811, 832-1119 residues. It consists of 3 chains, A (violet-red), B (dark khaki) and C (turquoise), which represent the subunits S1, S2 and S2' respectively.

MSA analysis among all three coronaviruses demonstrates that S protein of SARS-CoV-2 has a 77.71% sequence identity with Bat CoV and 77.14% identity with Human SARS **(Supplementary Figure S1A)**. All three S proteins are found to have a conserved C-terminal region. However, the N-terminal regions of S proteins display noteworthy differences. Given that there is significant sequence variation at S RBD in N-terminus, this might be the reason behind variation in its virulence and its receptor-mediated binding and entry into the host cell. According to our analysis of intrinsic disorder propensity, S protein of all three studied coronaviruses are highly structured as their predicted disorder propensity lies below 10% (**Table 1**). The mean PPID score of SARS-CoV-2, Human SARS, and Bat CoV is calculated

to be 1.41%, 1.12%, and 1.85% respectively. Graphs in **figures 3A, 3B** and **3C** represent the intrinsic disorder propensity of each residue in S protein of SARS-CoV-2, Human SARS and Bat CoV obtained from six disorder predictors.

Envelope (E) glycoprotein: E Glycoprotein is a multifunctional inner membrane protein that plays an important role in the assembly and morphogenesis of virions in the cell (Ruch & Machamer, 2012; Ujike & Taguchi, 2015) (DeDiego *et al*, 2007).
E protein consists of two ectodomains associated with N and C-terminal regions, and a transmembrane domain. It homo-oligomerize to form pentameric membrane destabilizing transmembrane (TM) hairpins to form a pore necessary for its ion channel activity (Torres *et al*, 2005). **Figure 4D** shows the NMR-structure (PDB ID: 2MM4) of Human SARS envelope glycoprotein of 8-65 residues (Li *et al*, 2014).

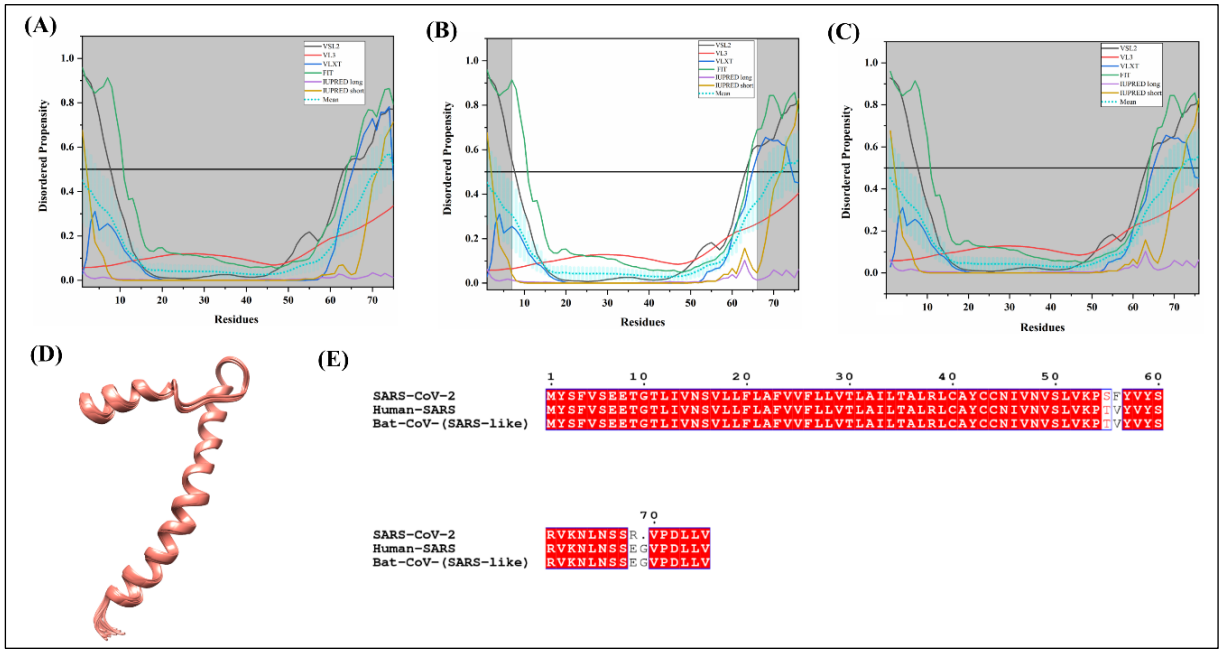

Figure 4. Analysis of intrinsic disorder propensity of Envelope glycoprotein. (A) Graph represents the IDP analysis in SARS-CoV-2**, (B)** Human SARS and **(C)** Bat CoV. The color schemes are same as figure 3. **(D)** The solution NMR generated structure of envelope glycoprotein of Human SARS (PDB ID: 2MM4). Consists of chain A (salmon colour) of 8-65 residues. **(E)** MSA profile of envelope glycoproteins of SARS-CoV-2, Human SARS and Red shaded sequences represent the similarity between the protein sequences.

MSA results illustrate (**Figure 4E**) that this protein is highly conserved, with only three amino acid substitutions in E protein of SARS-CoV-2 conferring its 96% sequence similarity with Human SARS and Bat CoV. Bat CoV shares 100% sequence identity with Human SARS. Mean PPID calculated for SARS-CoV-2, Human SARS, and Bat CoV E proteins are 5.33%, 6.58%, and 6.58% respectively (**Table 1**). The E protein is found to have a reasonably well-predicted structure. Our predictions suggest that residues of N- and C-terminals are displaying a higher tendency for the disorder. The last 18 hydrophilic residues (59-76) have been reported to adopt a random-coil conformation with and without the addition of lipid

membranes (Surya *et al*, 2013). Literature suggests that the last four amino acids of the C-terminal region of E protein containing a PZD-binding motif are involved in protein-protein interactions with a tight junction protein PALS1. PALS1 is involved in maintaining the polarity of epithelial cells in mammals (Teoh *et al*, 2010). We speculate that the disordered region content may be facilitating the interactions with other proteins as well. Respective graphs in **figure 4A, 4B,** and **4C** indicates the predicted intrinsic disorder in E proteins of SARS-CoV-2, Human SARS, and Bat CoV.

Membrane (M) glycoproteins: M Glycoprotein plays an important role in virion assembly by interacting with the N, and E proteins (Tseng *et al*, 2013, 2010; Corse & Machamer, 2003). Protein M interacts specifically with a short viral packaging signal containing coronavirus RNA in the absence of N protein highlights an important nucleocapsid-independent viral RNA packaging mechanism inside host cells (Narayanan *et al*, 2003). It gains high-mannose N-glycans in ER which is modified into complex N-glycans in the Golgi complex. Glycosylation of M protein is observed to be not essential for virion fusion in cell culture (Nal *et al*, 2005; Voss *et al*, 2009).

Cryo-EM and Tomography data indicate that M forms two distinct conformations: Compact M protein having high flexibility and low spike density; Elongated M protein having a rigid structure and narrow range of membrane curvature (Neuman *et al*, 2011). Regions of M glycoproteins are important as a dominant immunogen. This is reported as a deduced X-ray diffraction-based crystal structure (PDB ID: 3I6G) of HLA class I histocompatibility antigen, A-2 alpha chain, and Beta-2-microglobulin with small peptide of membrane glycoprotein (88-96 residues) complex in **figure 5D** (Liu *et al*, 2010).

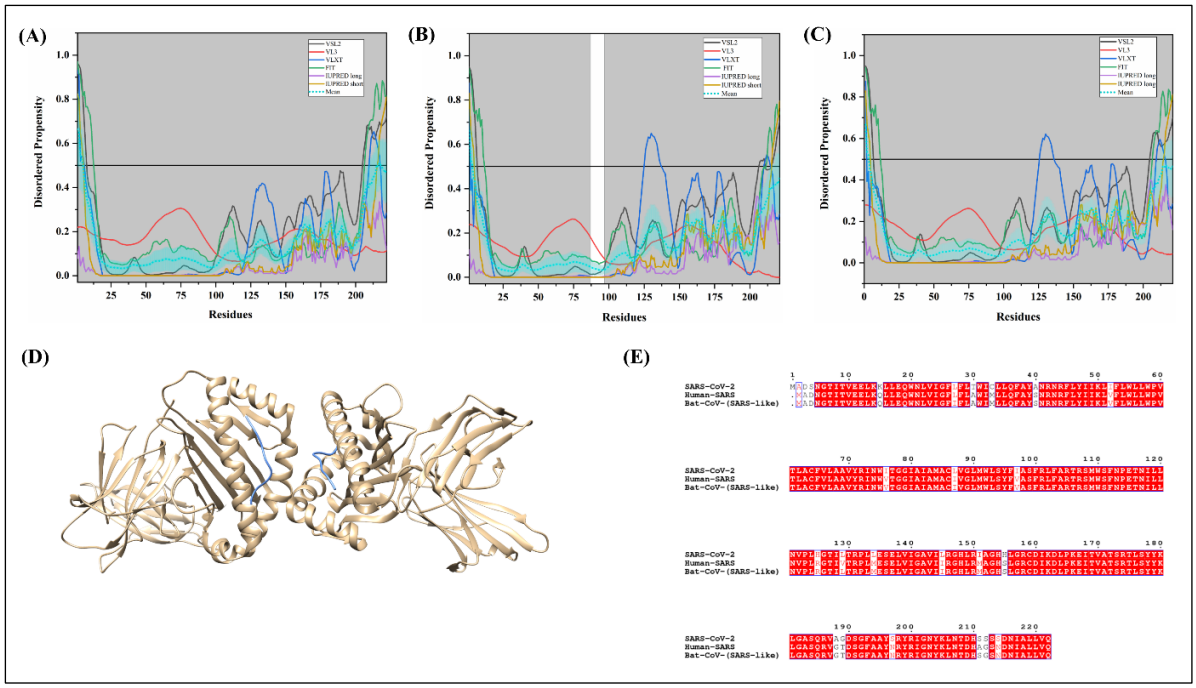

Figure 5. Analysis of intrinsic disorder propensity of Membrane glycoprotein. (A) Graph represents the IDP analysis in SARS-CoV-2**, (B)** Human SARS and **(C)** Bat CoV. The color schemes are same as figure 3. **(D)** A 2.20 Å resolution structure of complex HLA class I histocompatibility

antigen, A-2 alpha chain and Beta-2-microglobulin with immunogen peptide of membrane glycoprotein (9 amino acids) of Human SARS generated, by X-ray diffraction (PDB ID: 3I6G). Membrane glycoprotein consists of two chains, C and F chain (conifer blue colour) of 88-96 residues and other entities are represented in tan colour. **(E)** MSA profile of Membrane glycoproteins of SARS-CoV-2, Human SARS and Bat CoV (SARS-like). Red shaded sequences represent the similarity between the protein sequences.

M protein of SARS-CoV-2 has a sequence similarity of 90.05% with Bat CoV and 89.59% with Human SARS M proteins (**Figure 5E**). Our analysis revealed the intrinsic disorder in M proteins of SARS-COV-2 Human SARS, and Bat CoV as estimated mean PPID is 2.70%, 1.36%, and 1.36% respectively. This is in line with the previous publication by Goh *et al.* on Human SARS HKU4 where they found the mean PPID of 4% using additional predictors such as TopIDP and FoldIndex along with the predictors used in our study (Goh *et al*, 2013). **Figures 5A, 5B** and **5C** show the graphs for the predisposition of intrinsic disorder in residues of M proteins of SARS-CoV-2, Human SARS, and Bat CoV.

Nucleocapsid (N) protein: N protein is one of the major proteins playing a significant role in transcription, and virion assembly of coronaviruses (McBride *et al*, 2014). It binds to viral genomic RNA forming a ribonucleoprotein core required for encapsidation of RNA during viral particle assembly (Saikatendu *et al*, 2007). SARS-CoV virus-like particles (VLPs) formation has been reported to depend upon either M and E proteins or M and N proteins. For effective production and release of VLPs, co-expression of E and N proteins has been found to be necessary with M protein (Siu *et al*, 2008).

N protein of Human SARS consists of two structural domains: The N-terminal domain (NTD: 45-181 residues) and the C-terminal dimerization domain (CTD: 248-365 residues) with a disordered patch in between the domains. N protein has been demonstrated to bind viral RNA using both NTD and CTD (Chang *et al*, 2009). **Figure 6D1** displays the resolved crystal structure (PDB ID: 1SSK) of Human SARS nucleocapsid protein (45-181 residues) (Huang *et al*, 2004). **Figure 6D2** shows another X-ray diffraction generated crystal structure (PDB ID: 2GIB) of Human SARS nucleocapsid protein (270-366 residues) (Yu *et al*, 2006).

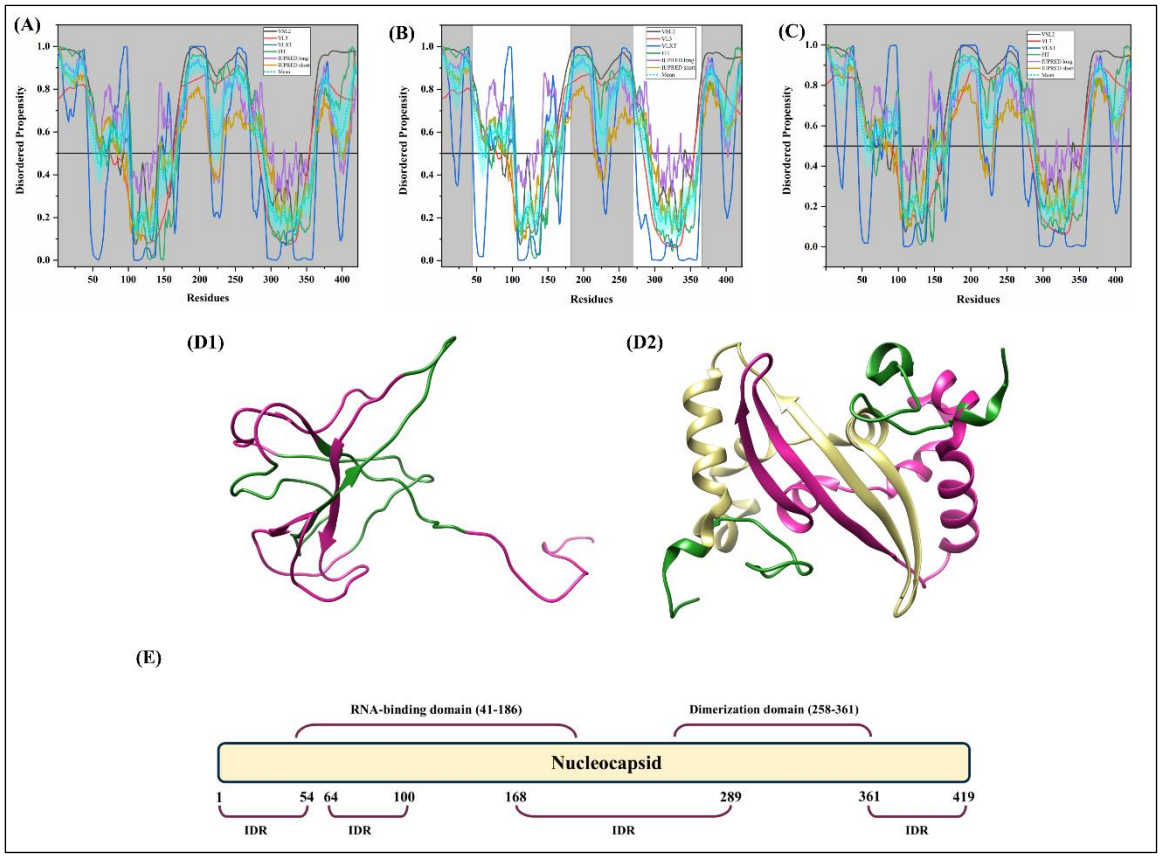

Figure 6. Analysis of intrinsic disorder propensity of Nucleocapsid. (A) Graph represents the IDP analysis in SARS-CoV-2**, (B)** Human SARS, and **(C)** Bat CoV. The color schemes are same as figure 3. **(D1)** Shows the PDB structure of the nucleocapsid of Human SARS generated by solution NMR (PDB ID: 1SSK). It Consists of only A chain of 45-181 residues (violet red colour). Residues: 1-57, 64-102 and 145-162 have disorder in their structure (forest green colour). **(D2)** X-ray diffraction generated 1.75Å resolution structure of nucleocapsid of Human SARS (PDB ID: 2GIB) of 270-366 residues. The structure is a homodimer of chains A (violet-red) and B (dark khaki). Residues 270-289 and 362-366 have shown disorder propensity in the available structures (forest green colour)**. (E)** Representation of domains of SARS-CoV-2 based on similarity with human SARS and predicted disordered regions of SARS-CoV-2.

419 amino acids long N protein of SARS-CoV-2 shows a percentage identity of 88.76% with N protein of Bat CoV N protein and 89.74% with Human SARS N protein (**Supplementary Figure S1B)**. Our analysis revealed the respective mean PPID of 60.38%, 71.09%, and 65.80% for SARS-COV-2 Human SARS, and Bat CoV N proteins. In accordance with the previously calculated intrinsic disorder (Goh *et al*, 2013), N protein is highly disordered in all three SARS viruses (**Table 1).** Graphs in **figure 6A, 6B,** and **6C** depicts the disorder in SARS-CoV-2 Human SARS, and Bat CoV nucleocapsid residues. that N and C-terminals are completely disordered with the central unstructured segment. As expected, N protein of novel coronavirus has similar residues showing the tendency for the disorder which is reported in Human SARS N protein prior to this study (Goh *et al*, 2013). The N and C-terminals are completely disordered with the central unstructured segment. In novel N protein, following residues 1-57, 64-102, 145-162, 166-289 and 362-422 are found to be disordered. The

residues are lying within the NTD and CTD region which due to their structural plasticity were not crystallized in Human SARS N protein. SARS-CoV-2 has a disordered segment from 168-289 residues while Human SARS has predicted to have an unstructured segment from 145-289 residues. Overall, all three N proteins are found to be highly disordered proteins.

Intrinsic disorder analysis of accessory proteins of Coronaviruses:

Literature suggests that some proteins are translated from the genes interspersed in between the genes of structural proteins. These proteins are termed as accessory proteins, and many of them are proposed to work in viral pathogenesis (Narayanan *et al*, 2008).

ORF3a *and* ***ORF3b*****:** ORF3a is, of molecular weight ~31 kDa, a multifunctional protein that has been found localized in different organelles inside host cells. Also referred to as U274, X1, and ORF3, the gene for this protein is present between S and E genes of the SARS-CoV genome (Yu *et al*, 2004; Yuan *et al*, 2005a; Tan *et al*, 2004). The homo-tetramer complex of ORF3a has been demonstrated to form a potassium-ion channel on the host cell plasma membrane (Lu *et al*, 2006). It performs a major function during virion assembly by co-localizing with E, M, and S viral proteins (Tan, 2005; McBride & Fielding, 2012). ORF3b protein localizes in the cytoplasm, nucleolus, and outer membrane of mitochondria (Yuan *et al*, 2006, 2005c). In Huh 7 cells, its over-expression has been linked with the activation of AP-1 via ERK and JNK pathways (Varshney & Lal, 2011). Transfection of ORF3b-EGFP leads to cell growth arrest in the G0/G1 phase of Vero, 293 and COS-7 cells (Yuan *et al*, 2005b). ORF3a induces apoptosis via caspase 8/9 directed mitochondrial-mediated pathways while ORF3b is reported to affect only caspase 3 pathways (Khan *et al*, 2006; Law *et al*, 2005b).

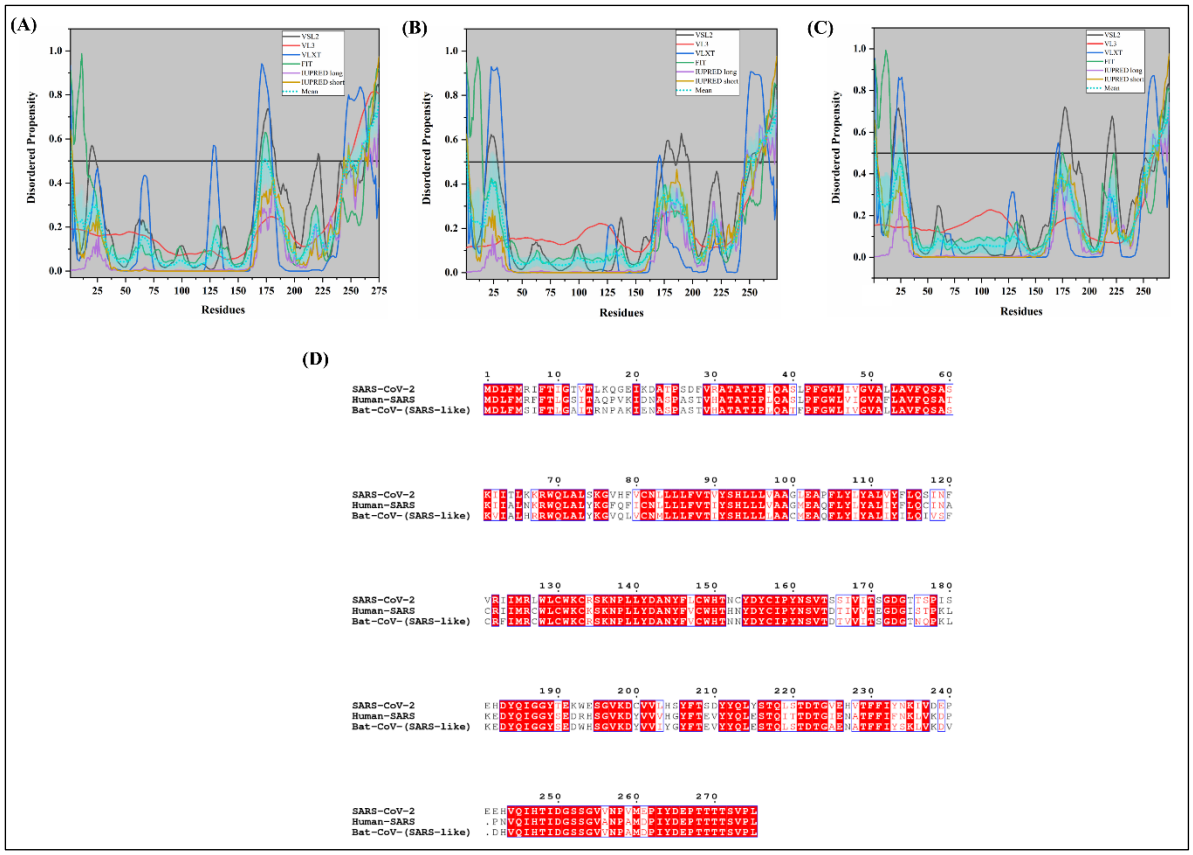

Figure 7. Analysis of intrinsic disorder propensity of ORF 3a protein. (A) The graph represents the IDP analysis in SARS-CoV-2 of ORF 3a protein**, (B)** Human SARS of ORF 3a protein and **(C)** Bat CoV of ORF 3 protein. The color schemes are same as figure 3. **(D)** MSA profile of ORF 3a of SARS-CoV-2, Human SARS, and ORF 3 of Bat CoV. Red shaded sequences represent the similarity between the protein sequences.

On performing MSA, shown in **figure 7D**, we found that ORF 3a protein of SARS-COV-2 is found to be closer to ORF3a of Bat CoV (73.36%) than ORF3a of Human SARS (72.99%).

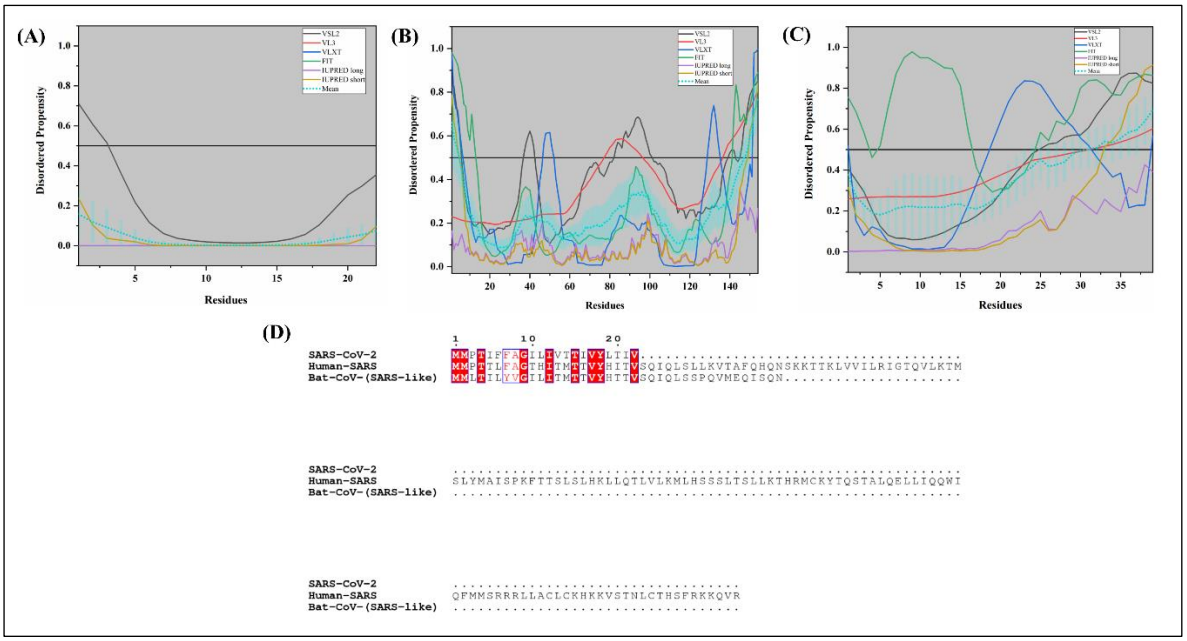

Figure 8. Analysis of intrinsic disorder propensity of ORF 3b protein of Human SARS CoV. (A) The graph represents the IDP analysis in SARS-CoV-2**, (B)** Human SARS, and **(C)** Bat CoV. The color schemes are same as figure 3. **(D)** MSA profile of ORF 3b of SARS-CoV-2, Human SARS, and Bat CoV. Red shaded sequences represent the similarity between the protein sequences.

Graphs in **Figures 7A, 7B** and **7C** depict the propensity of disorder in ORF 3a proteins of novel SARS-CoV-2, Human SARS, and Bat CoV (SARS-like). Mean PPIDs in ORF 3a proteins are: SARS-CoV-2 - 9.09%, Human SARS - 8.76%, and Bat CoV (SARS-like) - 6.20%. According to the analysis of intrinsic disorder, the mean PPID in ORF3b proteins of SARS-CoV-2, Human SARS, and Bat CoV are 0%, 7.14%, and 23.08% respectively, as represented in **figures 8A, 8B,** and **8C.** MSA results demonstrate that (**Figure 8D**), ORF3b of SARS-CoV-2 is not closer to ORF3b protein of Human SARS and to ORF3b protein of Bat-CoV, having a sequence similarity of only 54.55% and 59.09% respectively.

ORF6: ORF6 is a very short SARS coronavirus protein having ~63 residues. Also known as P6, is a membrane-associated protein that serves as an interferon (IFN) antagonist (Kopecky-Bromberg *et al*, 2007). It downregulates the IFN pathway by blocking a nuclear import protein, karyopherin α2. Using its C-terminal residues, ORF6 disrupts karyopherin import complex in the cytosol and, therefore, hampers the movement of transcription factors like STAT1 into the nucleus (Kopecky-Bromberg *et al*, 2007; Frieman *et al*, 2007). It contains a YSEL motif near its C-terminal region, which functions in protein internalization from the plasma membrane into the endosomal vesicles (Netland *et al*, 2007). Another study has also demonstrated the presence of ORF6 in endosomal/lysosomal compartments (Netland *et al*, 2007; Gunalan *et al*, 2011).

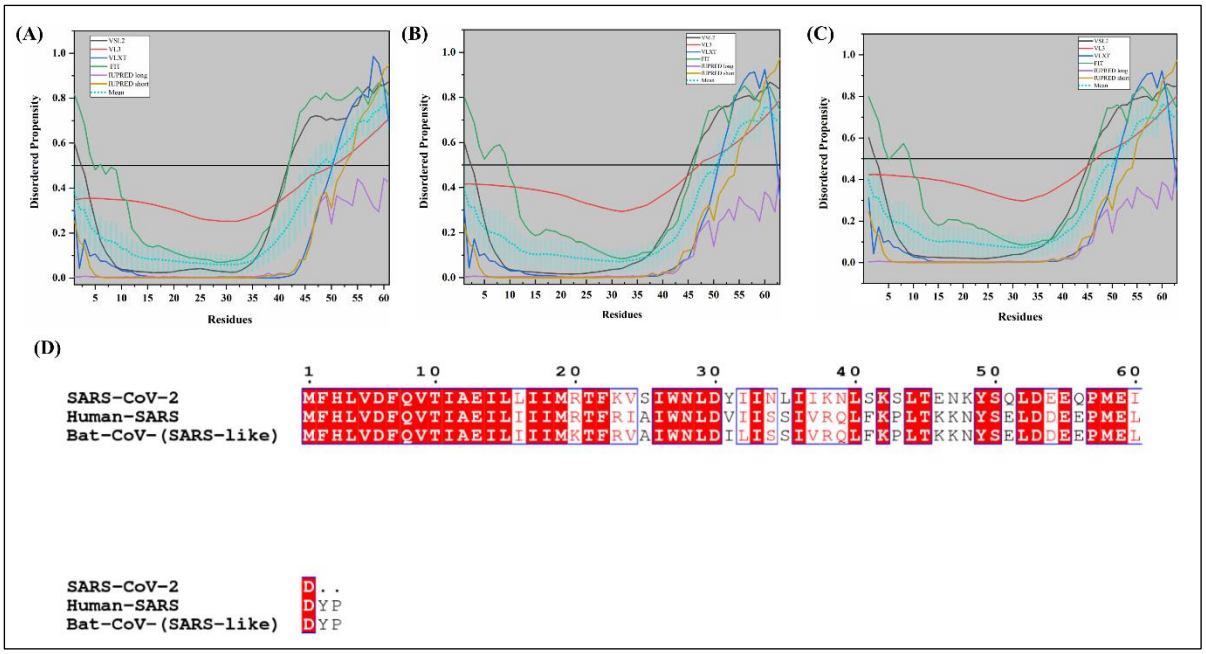

Figure 9. Analysis of intrinsic disorder propensity of ORF 6 protein. (A) Graphs represent the IDP analysis in SARS-CoV-2**, (B)** Human SARS, and **(C)** Bat CoV. The color schemes are same as figure 3. **(D)** MSA profile of ORF 6 of SARS-CoV-2, Human SARS and Bat CoV. Red shaded sequences represent the similarity between the protein sequences.

MSA results demonstrate that (**Figure 9D**), SARS-CoV-2 ORF6 is closer to ORF6 protein of Human SARS, having a sequence similarity of 68.85% than ORF6 of Bat CoV (SARS-like) (67.21%). Novel SARS-CoV-2 ORF6 is calculated to be the second most disordered structural protein with a PPID of 22.95%, especially, at the C-terminal region.

Our analysis of intrinsic disorder using six predictors have revealed the mean PPID in ORF6 proteins of SARS-CoV-2, Human SARS, and Bat CoV are analyzed to be 22.95%, 20.63%, and 20.63% respectively (**Table 1**). Graphs in **figures 9A, 9B** and **9C** illustrate the disorder for each residue in ORF6 proteins of all three studied coronaviruses. Moderately disordered protein ORF6 is predicted to have an intrinsic disorder near its C-terminal region **(Figures 9A, B, C)**. These regions are significant for the biological activities of ORF6. As mentioned above, this hydrophilic region contains lysosomal targeting motif (YSEL) and diacidic motif (DDEE) responsible binding and recognition during translocation (Netland *et al*, 2007). However, the N-terminal region is not showing any disordered prediction. The literature on Human SARS suggests that the 1-38 amino acid N-terminal region is alpha-helical and embedded in the membrane but not a transmembrane protein (Zhou *et al*, 2010).

ORF7a** and **ORF7b: Alternatively called U122, ORF7a is a type I transmembrane protein (Fielding *et al*, 2004; Huang *et al*, 2006). It has been proven to localize in ER, Golgi, and peri-nuclear space. The presence of a KRKTE motif near the C-terminal imports the protein from the ER to the Golgi apparatus (Fielding *et al*, 2004; Huang *et al*, 2006). It contributes to viral pathogenesis by activating the release of pro-inflammatory cytokines and chemokines like IL-8 and RANTES (Kanzawa *et al*, 2006; Law *et al*, 2005a). In another study,

overexpression of BCL-XL in 293T cells blocked the ORF7a mediated apoptosis (Tan *et al*, 2007). **Figure 10D** represents the 1.8Å X-ray diffraction-based crystal structure (PDB ID: 1XAK) of ORF7a on Human SARS revealed its compact seven-stranded β topology similar to Ig-superfamily members (Nelson *et al*, 2005). Its structure includes a signal peptide, a luminal domain, a transmembrane domain and a short cytoplasmic tail at 5' end (Nelson *et al*, 2005; McBride & Fielding, 2012).

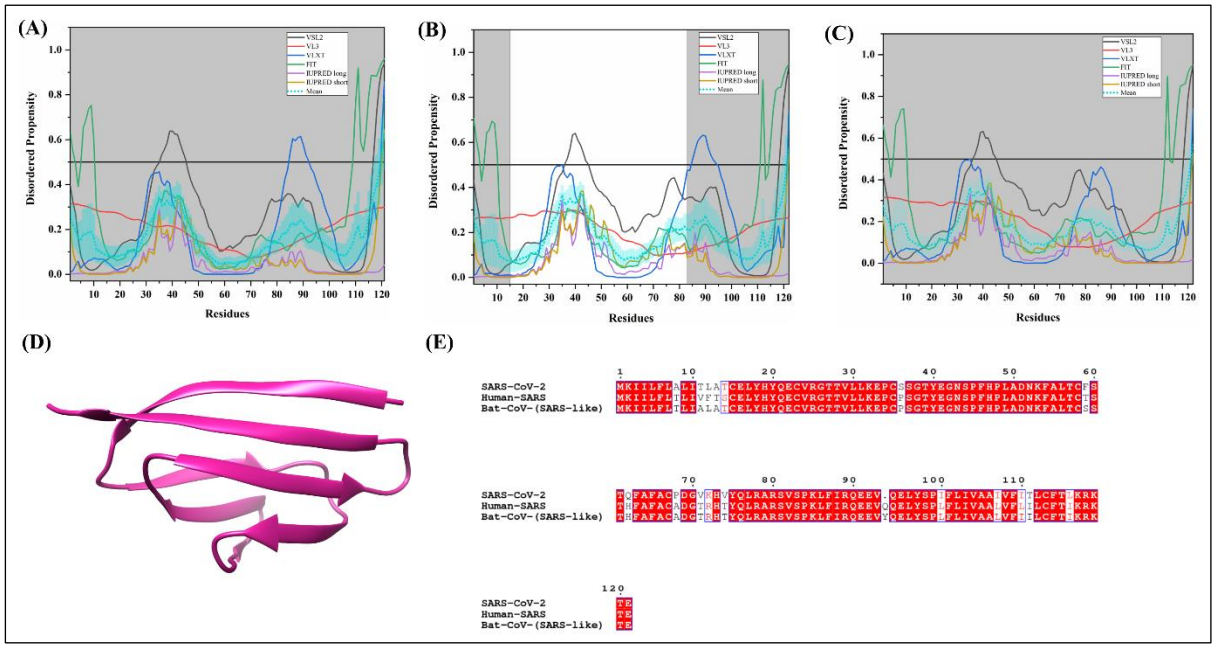

Figure 10. Analysis of intrinsic disorder propensity of ORF 7a protein. (A) The graph represents the IDP analysis in SARS-CoV-2**, (B)** Human SARS, and **(C)** Bat CoV. The color schemes are same as figure 3. **(D)** A 1.8Å resolution PDB structure of ORF 7a of Human SARS (PDB ID: 1XAK) generated by X-ray diffraction. It consists of only A chain of 14-96 residues (violet red colour). **(E)** MSA profile of ORF7a of SARS-CoV-2, Human SARS and Bat CoV. Red shaded sequences represent the similarity between the protein sequences.

We found that 121 residues long ORF7a protein of SARS-CoV-2 shares 89.26% and 85.95% sequence identity with ORF7a proteins of Bat CoV, and Human SARS (**Figure 10E**). ORF7b protein of SARS-CoV-2 is found to be closer with Human SARS (81.40%) than Bat sequence with 79.07% sequence identity **figure 11D**.

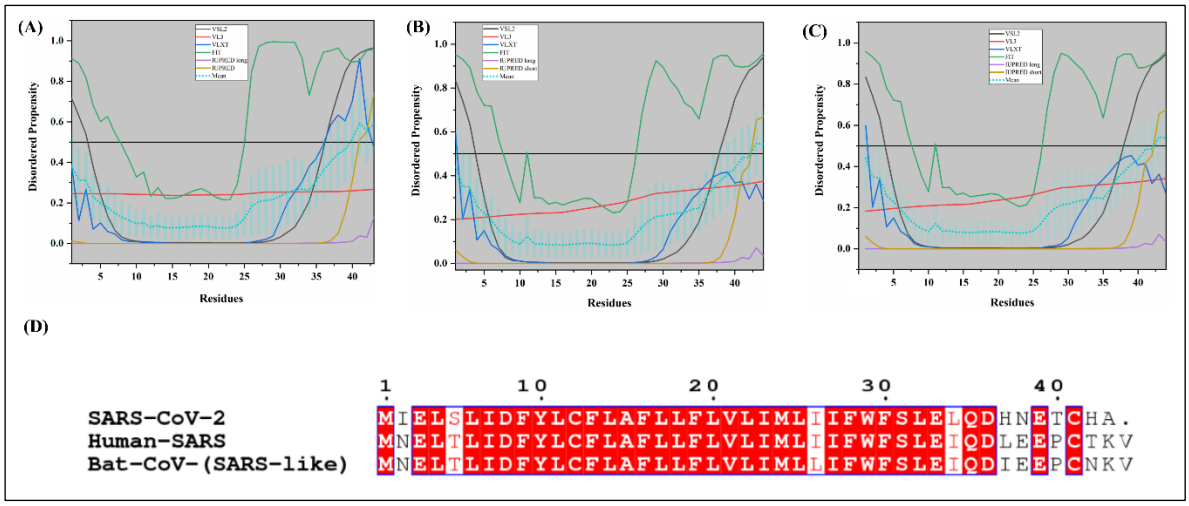

Figure 11. Analysis of intrinsic disorder propensity of ORF 7b protein. (A) The graph represents the IDP analysis in SARS-CoV-2**, (B)** Human SARS and **(C)** Bat CoV. The color schemes are same as figure 3. **(D)** MSA profile of ORF7b of SARS CoV-2, Human SARS, and Bat CoV. Red shaded sequences represent the similarity between the protein sequences.

As can be observed from **table 1,** our disorder prediction resulted in the overall PPID for ORF7a proteins are 1.65% for SARS-CoV-2, 0.82% for Bat CoV and 0.82% for Human SARS. Mean PPIDs estimated for ORF7b proteins are: 9.30% of novel CoV, 4.55% for Bat CoV and 4.55% Human SARS. Graphs in **figures 10A, 10B, and 10C** represent the predisposed residues for disorder in ORF7a proteins of SARS-CoV-2, Human SARS, and Bat CoV respectively. Graphs in **figures 11A, 11B, and 11C,** depict the predisposed residues for disorder in ORF7b proteins of SARS-CoV-2, Human SARS, and Bat CoV respectively. According to our analysis, both proteins in all three studied coronaviruses have a significantly ordered structure.

ORF7b is an integral membrane protein that has been shown to localize in the Golgi complex (Schaecher *et al*, 2007; Kopecky-Bromberg *et al*, 2006). The same reports also confirm the role of ORF7b as an accessory as well as a structural protein for SARS-CoV virion.

ORF8a and ORF8b: In animals and in isolates from early human infections, the ORF8 gene codes for a single ORF8 protein. However, it was observed that in late infections, at middle and late stages, a 29 nucleotide deletion led to the formation of two distinct proteins; ORF8a and ORF8b having 39 and 84 residues respectively (Oostra *et al*, 2007a; Chinese SARS Molecular Epidemiology Consortium, 2004). Both proteins have a distinct conformation than the longer ORF8 protein. It has been reported that overexpression of ORF8b resulted in the downregulation of E protein while the proteins ORF8a and ORF8/ORF8ab have no effect on the expression of protein E. Also, ORF8/ORF8ab was found to interact very strongly with proteins S, ORF3a and ORF7a. ORF8a interacts with S and E proteins, whereas ORF8b protein interacts with E, M, ORF3a and ORF7a proteins (Keng *et al*, 2006).

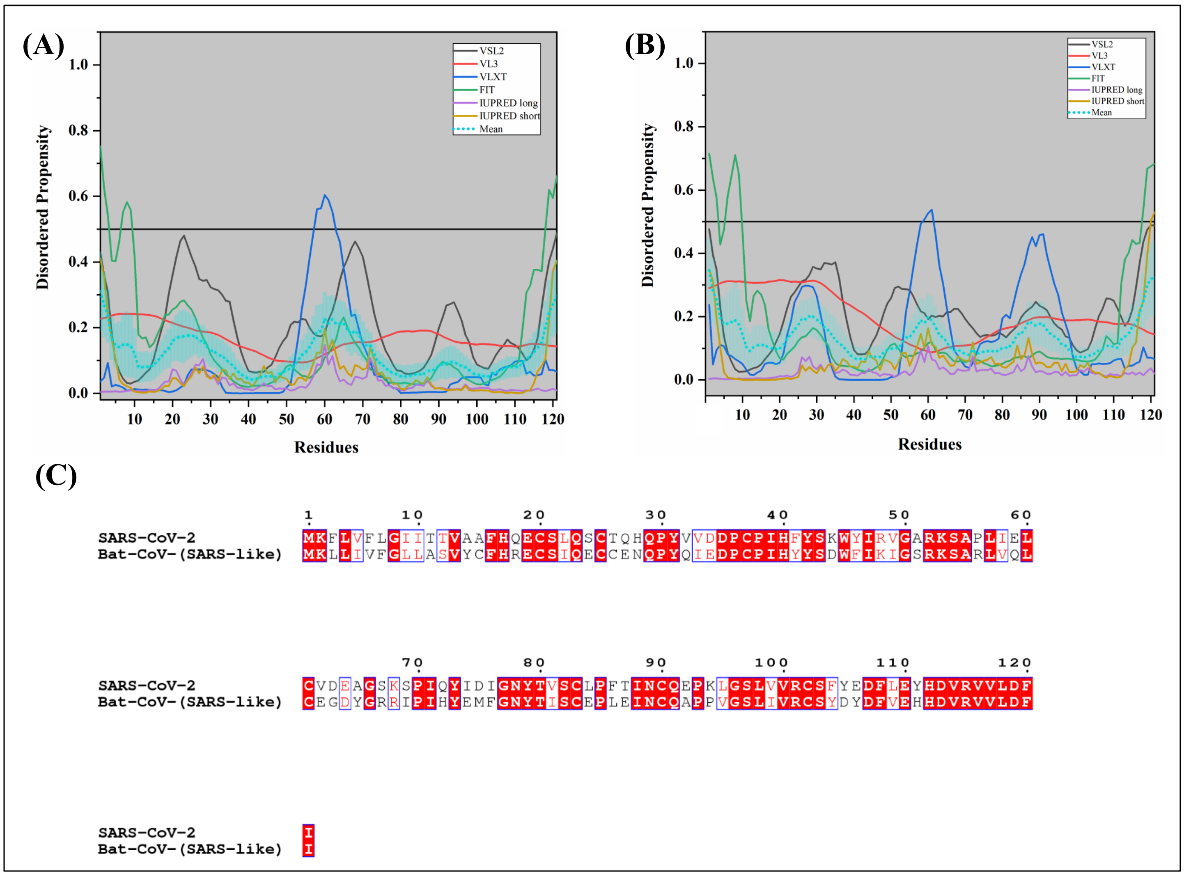

Figure 12. Analysis of intrinsic disorder propensity of ORF 8 protein. (A) Graph represent the IDP analysis in SARS-CoV-2 and **(B)** Bat CoV. The color schemes are same as figure 3. **(C)** MSA profile of ORF 8 of SARS-CoV-2 and Bat CoV. Red shaded sequences represent the similarity between the protein sequences.

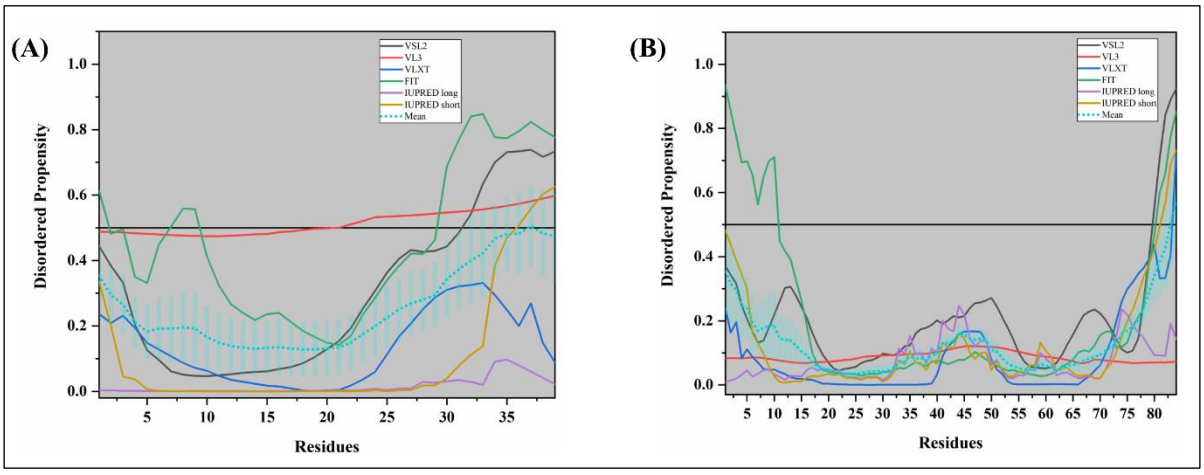

Figure 13. Analysis of intrinsic disorder propensity of ORF 8a and ORF 8b protein of Human SARS. (A) The graph represents the IDP analysis in Human SARS of ORF 8a protein. **(B)** The graph represents the IDP analysis in Human SARS of ORF 8b protein. The color schemes are same as figure 3.

Early SARS CoV-2 isolates also have a single and longer ORF8 protein having 121 residues and according to our analysis, it shares a 90.05% sequence identity with ORF8 protein of Bat CoV (**Figure 12C**). Our IDP analysis revealed no disorder in both ORF8 proteins (of SARS-CoV-2, Bat CoV) as can be observed from the graphs in **figures 12A** and **12B**. Both proteins are analyzed to be completely structured having a mean PPID of 0.00%. In ORF8a and ORF8b proteins of the Human SARS, the predicted disorder is estimated to be 2.56%, and 2.38% respectively **(Table 1)**. Graphs in (**Figures 13A** and **13B**) illustrate a little disorder near the end terminals of ORF8a and ORF8b proteins.

ORF9b: This protein is expressed from an alternative ORF within the N gene through a leaky ribosome binding process (Xu *et al*, 2009). Inside host cells, ORF9b enters the nucleus which is a cell cycle independent process and passive entry. This has been shown to interact with a nuclear export protein receptor Exportin 1 (Crm1) using which it translocate out of the nucleus (Sharma *et al*, 2011). A 2.8Å resolved crystal structure (PDB ID: 2CME) of ORF9b protein of Human SARS exposed the presence of a dimeric tent-like β structure along with central hydrophobic amino acids (**Figure 14D)**. The published structure has highly polarized distribution of charges with positively charged residues on one side and negatively charged on the other (Meier *et al*, 2006).

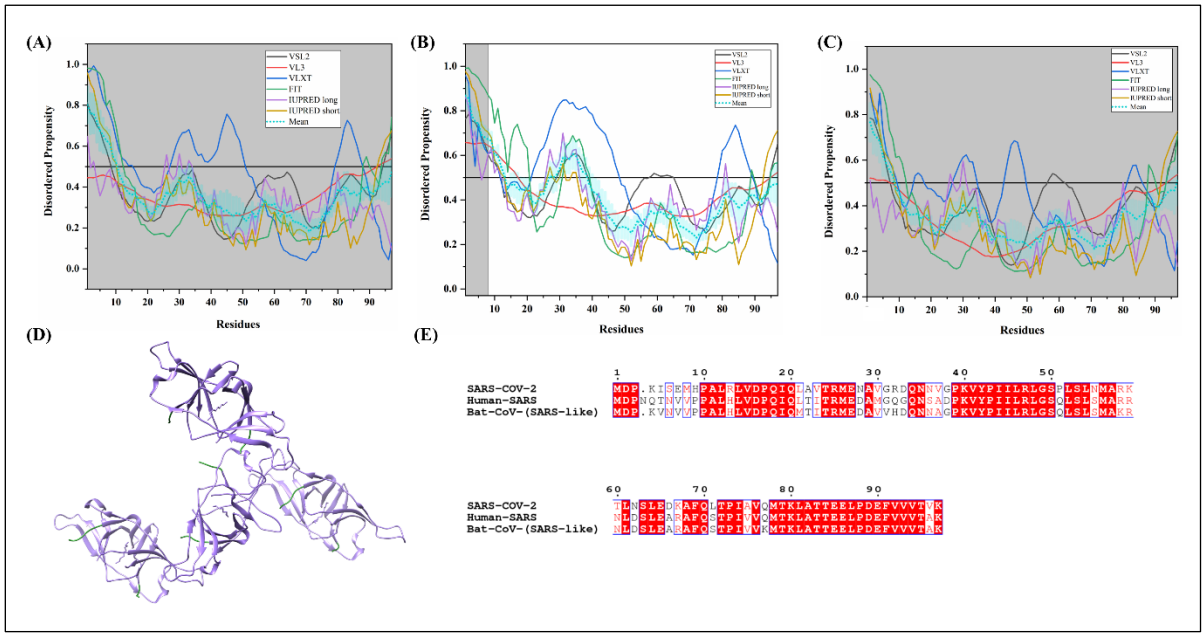

Figure 14. Analysis of intrinsic disorder propensity of ORF 9b protein. (A) The graph represents the IDP analysis of ORF9b of SARS-CoV-2, **(B)** Human SARS and **(C)** Bat CoV of ORF 9 protein. The color schemes are same as figure 3. **(D)** A 2.8Å resolution PDB structure of ORF 9b of Human SARS generated by X-ray diffraction (PDB ID: 2CME). Consists of A, B, E, G chains of 9-98 residues (purple colour); C, D, F, H chains of 10-98 residues (purple colour). **(E)** MSA of ORF9b of Human SARS and ORF9 of Bat CoV. Red shaded sequences represent the similarity between the protein sequences.

Based on the sequence availability of Accession ID NC_045512.2, the translated protein sequence of ORF9b is not reported for novel SARS-CoV-2 yet. However, based on the report by Wu and colleagues, they have already annotated the sequences of SARS-CoV-2, so we have taken the sequences and analyzed the intrinsic disorder in our manuscript. According to the MSA represented in **figure 14E**, ORF9b protein of SARS-CoV-2 shares 73.20% similarity with Human SARS and 74.23% similarity with Bat CoV.

Our IDP analysis (**Table 1**) shows that ORF9b of Human SARS is a moderately unstructured protein with a mean PPID estimated to 26.53%. As depicted in the graphs of **figure 14A, 14B, and 14C** disorder mainly lies near N-terminal end 1-10 residues and 28-40 residues near the central region with a well-ordered inner core of Human SARS ORF9b protein. X-ray diffraction-based crystal structure of ORF9b has a missing electron density of the first 8 residues and 26-37 residues near the central region. This might be because disordered regions are difficult to crystallize due to their highly dynamic structural conformations. SARS-CoV-2 ORF9b protein with a mean PPID of 10.31% has a N-terminal (1-10 residues) predicted disordered segment. ORF9b of Bat coronavirus is analyzed to have an intrinsic disorder of 9.28%, comparatively lesser than Human SARS ORF9b protein.

ORF10: The newly emerged SARS-CoV-2 has ORF10 protein of 38 amino acids. ORF10 of SARS-CoV-2 has a 100% sequence similarity with Bat CoV strain Bat-SL-CoVZC45 (Wu *et al*, 2020b). However, we have not done the IDPs analysis for ORF10 from the Bat-SL-CoVZC45 strain since we have taken different strain of Bat CoV (reviewed strain HKU3-1) in our study. Thus, we have done the IDP analysis for ORF10 protein of SARS-CoV-2 only, according to which, mean PPID is calculated to be 0.00%. **Figure 15** shows the graph of the predisposition of intrinsic disorder in residues of ORF10 protein.

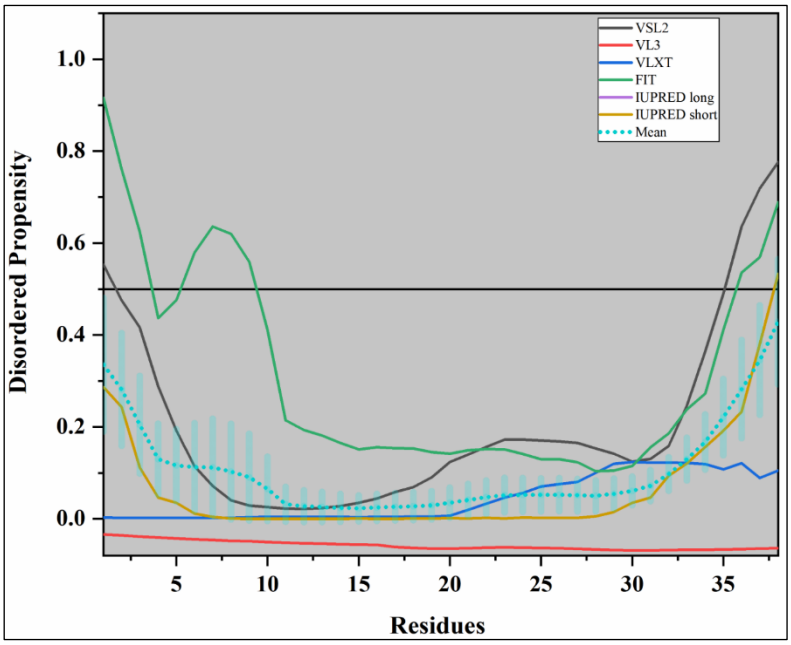

Figure 15. Analysis of intrinsic disorder propensity of ORF10 protein of SARS-CoV-2. The color schemes are same as above.

ORF14: This is a 70 amino acids long uncharacterized protein of unknown function is present in Human SARS and Bat CoV. However, SARS-CoV-2 ORF14 is of 73 amino acid long protein. According to the MSA, ORF14 of SARS-CoV-2 have 77.14% similarity with Human-SARS and 72.86% similarity with Bat CoV as represented in **figure 16D**. We have performed the IDP analysis to understand the presence of disorder in this protein. The graph in **figure 16A, 16B and 16C** illustrates the intrinsic disorder in residues of ORF14 of SARS-CoV-2, Human-SARS, and Bat CoV; having calculated mean PPID is to be 0.00%, 2.86%, and 0.00% respectively.

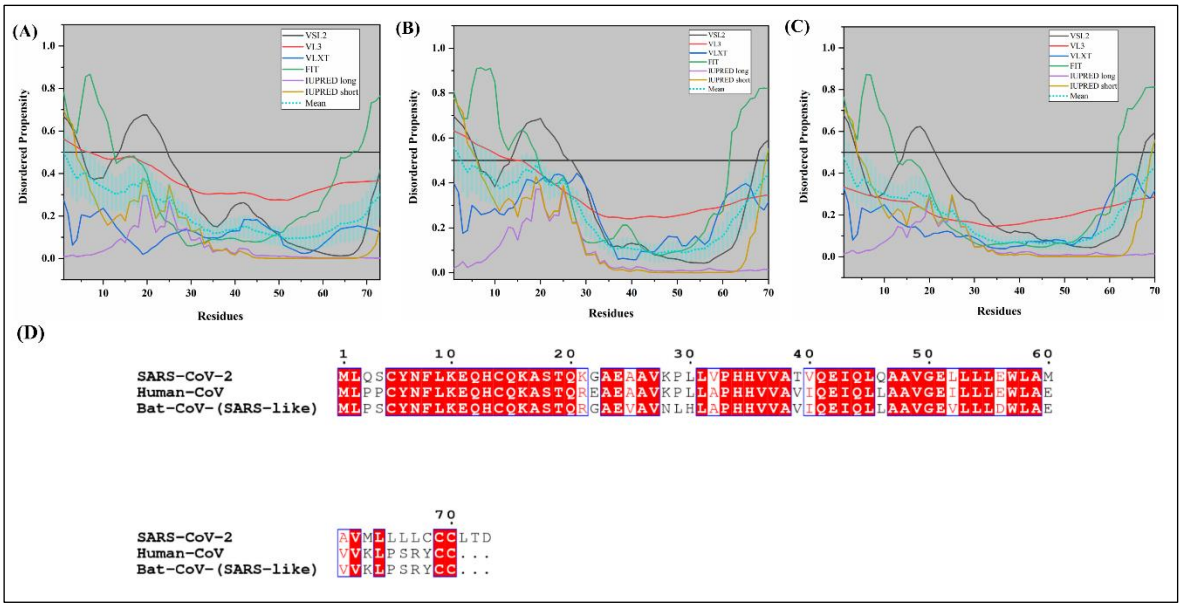

Figure 16. Analysis of intrinsic disorder propensity of ORF 14 protein. (A) The graph represents the IDP analysis of SARS-CoV-2, **(B)** Human-SARS and **(C)** Bat CoV of ORF 14 protein. The color schemes are same as figure 3. **(D)** MSA of ORF 14 of SARS-CoV-2, Human SARS and Bat CoV. Red shaded sequences represent the similarity between the protein sequences.

Intrinsic disorder analysis of non-structural proteins (Nsps) of Coronaviruses:

In coronaviruses, due to ribosomal leakage during translation two-third of the RNA genome is processed into two polyproteins: (i) Replicase polyprotein 1a and (ii) Replicase polyprotein 1ab. Both contain non-structural proteins (Nsp1-10) in addition to different proteins required for viral replication and pathogenesis. Replicase polyprotein 1a contains an additional Nsp11 protein of 13 amino acids function of which is not investigated yet. The longer replicase polyprotein 1ab of 7073 amino acids accommodates five other non-structural proteins (Nsp12-16) (Thiel *et al*, 2003). These proteins assist in ER membrane-induced vesicle formation, which acts as sites for replication and transcription. In addition to this, non-structural proteins work as proteases, helicases, and mRNA capping and methylation

enzymes, crucial for virus survival and replication inside host cells (Thiel *et al*, 2003; Fan *et al*, 2004).

Table 2: Evaluation of mean predicted percentage disorder in non-structural proteins of novel SARS-CoV-2, Human SARS, and Bat CoV.

Proteins	**SARS-CoV-2**		**Human SARS** (UniProt ID: P0C6X7)		**Bat CoV (SARS-like)** (UniProt ID: P0C6W2)	
	Length of protein (Residues) (NCBI RefSeq accession ID)	**Mean PPID**	**Length of protein (Residues)**	**Mean PPID**	**Length of protein (Residues)**	**Mean PPID**
Nsp1 (Host translation inhibitor)	180 (1-180) (YP_009725297.1)	12.78	180 (1 – 180)	14.44	179 (1-179)	12.85
Nsp2	638 (181-818) (YP_009725298.1)	5.17	638 (181 – 818)	2.04	639 (180-818)	2.03
Nsp3 (papain-like proteases, 3CL-Pro)	1945 (819-2763) (YP_009725299.1	7.40	1922 (819 – 2740)	7.91	1916 (819 – 2734)	7.78
Nsp4	500 (2764-3263) (YP_009725300.1	0.80	500 (2741 – 3240)	0.60	500 (2735 – 3234)	0.60
Nsp5 (Protease, 3CL-Pro)	306 (3264-3569) (YP_009725301.1	1.96	306 (3241 – 3546)	1.96	306 (3235 – 3540)	1.96
Nsp6	290 (3570-3859) (YP_009725302.1	1.03	290 (3547 – 3836)	1.03	290 (3541 – 3830)	4.48

Nsp7 (multimeric RNA polymerase)	83 (3860-3942) (YP_009725303.1)	9.64	83 (3837 – 3919)	9.64	83 (3831 – 3913)	9.64
Nsp8 (multimeric RNA polymerase)	198 (3943-4140) (YP_009725304.1)	23.74	198 (3920 – 4117)	23.74	198 (3914 – 4111)	22.22
Nsp9	113 (4141-4253) (YP_009725305.1)	7.08	113 (4118 – 4230)	7.96	113 (4112 – 4224)	7.08
Nsp10	139 (4254-4392) (YP_009725306.1)	5.04	139 (4231 – 4369)	5.04	139 (4225 – 4363)	5.04
Nsp12 (RNA-directed RNA polymerase, RDRP)	932 (4393-5324) (YP_009725307.1)	0.43	932 (4370 – 5301)	0.43	932 (4364 – 5295)	0.43
Nsp13 (Viral helicase)	601 (5325-5925) (YP_009725308.1)	0.67	601 (5302 – 5902)	0.67	601 (5296 – 5896)	0.67
Nsp14 (Guanine-N7 methyltransferase/ExoN-nsp14)	527 (5926-6452) (YP_009725309.1)	0.38	527 (5903 – 6429)	0.38	527 (5897 – 6423)	0.57
Nsp15 (Uridylate-specific endoribonuclease/ NendoU)	346 (6453-6798) (YP_009725310.1)	1.73	346 (6430 – 6775)	2.60	346 (6424 – 6769)	2.60
Nsp16 (2'-O-methyltransferase)	298 (6799-7096) (YP_009725311.1)	5.37	298 (6776 – 7073)	3.02	298 (6770 – 7067)	3.02

Global Analysis of Intrinsic Disorder in Replicase polyprotein 1ab:

Table 2 represents the PPID mean scores of the 15 non-structural proteins of Replicase polyprotein 1ab in SARS-CoV-2, Human SARS and Bat CoV that were obtained by combining the results from six disorder predictors (**Supplementary Table 4-6**). **Figures 17A, 17B and 17C** show the 2D-disorder plots of the Nsps coded by ORF1ab in SARS-CoV-2, Human SARS and Bat CoV respectively. Based on the PPID Mean scores in **table 2**, **figures 17A, 17B, 17C** and PPID based classification (Rajagopalan *et al*, 2011), we conclude that none of the Nsps in SARS-CoV-2, Human SARS and Bat CoV are highly disordered. The highest disorder was observed in Nsp8 of all three Coronaviruses. Both Nsp1 and Nsp8 are moderately disordered proteins (10% ≤ PPID ≤ 30%). We also observed that Nsp2, Nsp3, Nsp5, Nsp6, Nsp7, Nsp9, Nsp10, Nsp15 and Nsp16 have disorder less than 10% and hence, belong to the category of highly ordered proteins. The other non-structural proteins, namely, Nsp4, Nsp12, Nsp13 and Nsp14 have negligible levels of disorder (PPID < 1%) which tells us that these are highly structured proteins.

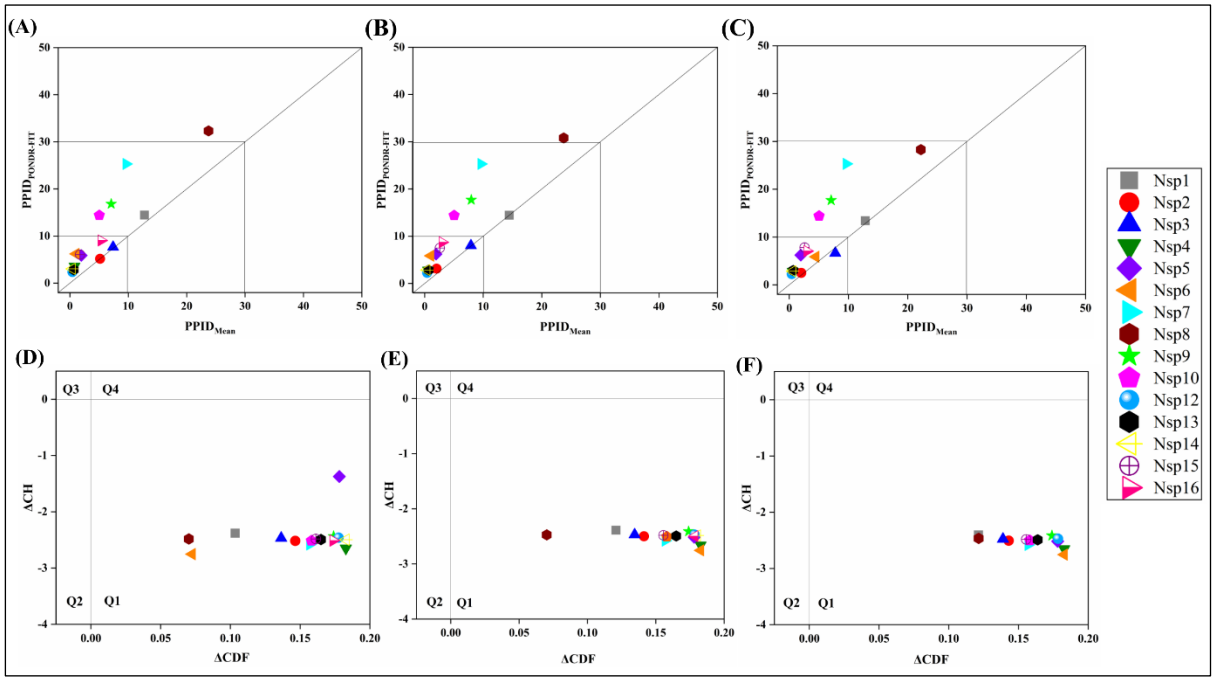

Figure 17. Analysis of overall disorder status of non-structural proteins (Nsps) coded by ORF1ab in SARS-CoV-2, Human SARS CoV and Bat CoV: 2D plot representing $PPID_{PONDR\text{-}FIT}$ vs $PPID_{Mean}$ in **(A)** SARS-CoV-2 **(B)** Human SARS and **(C)** Bat CoV (SARS-like). In CH-CDF plot of the proteins of **(D)** SARS-CoV-2 **(E)** Human SARS and **(F)** Bat CoV, the Y coordinate of each protein spot signifies distance of corresponding protein from the boundary in CH plot and the X coordinate value corresponds to the average distance of the CDF curve for the respective protein from the CDF boundary.

The CH-CDF analysis of the Nsps from SARS-CoV-2, Human SARS and Bat CoV have been represented in figures 17D, 17E and 17F respectively. It was observed that all the nsps of the three coronaviruses occur in quadrant Q1, which helps conclude that all the nsps are predicted to be ordered according to the binary predictors, CH and CDF.

Replicase polyprotein 1ab: The longer replicase polyprotein 1ab is 7073 amino acids which contain 15 non-structural proteins that are mentioned in **table 2**. Nsp1, Nsp2, and Nsp3 are cleaved using a viral papain-like proteinase (Nsp3/PL-Pro) while the rest are cleaved by another viral 3C-like proteinase (Nsp5/3CL-Pro). Based on the cleavage of replicase 1ab polyprotein of Human SARS by two proteases, we have shown disorder propensity at the cleavage sites with few residues spanning the terminals (**Figure 18**). Interestingly, we observed that all the cleavage sites are largely disordered, suggesting that intrinsic disorder may have a role in maturation of individual non-structural proteins. As the percentage identities of Nsps of Human SARS are closer with Nsps of SARS-CoV-2, we speculate the of presence of intrinsic disorder at the cleavage sites in polyproteins of SARS-CoV-2. Additionally, their structural and functional properties are thoroughly described below with the predicted intrinsic disorder regions.

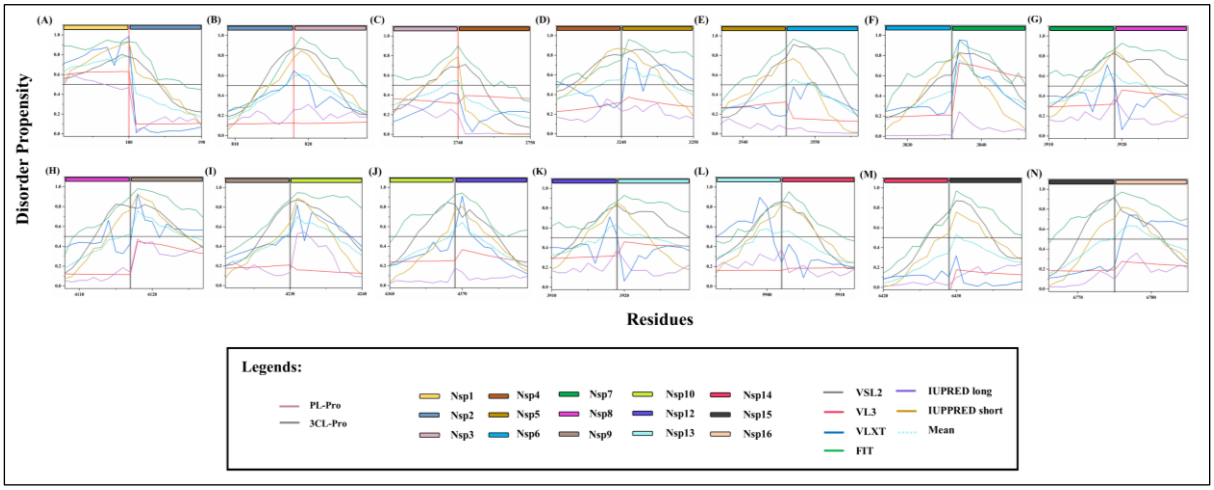

Figure 18. **Intrinsic disorder at cleavage sites of replicase 1ab polyprotein of Human SARS**. Plots A to N denote the cleavage sites (magenta coloured bar for PL-Pro protease and grey coloured bar for 3CL-Pro protease) in relation to disordered regions present between the individual proteins (Nsp1-16) of replicase 1ab polyprotein of Human SARS. All proteins are represented by different coloured horizontal bars.

Non-structural protein 1 (Nsp1): This works as a host translation inhibitor as it binds to the 40S subunit of the ribosome and blocks the translation of cap-dependent mRNAs as well as mRNAs that uses internal ribosome entry site (IRES) (Lokugamage *et al*, 2012). **Figure 19D** shows the NMR solution structure (PDB ID: 2GDT) of Human SARS nsp1 protein (13-128 residues), residues 117-180 are missing from this structure (Almeida *et al*, 2007).

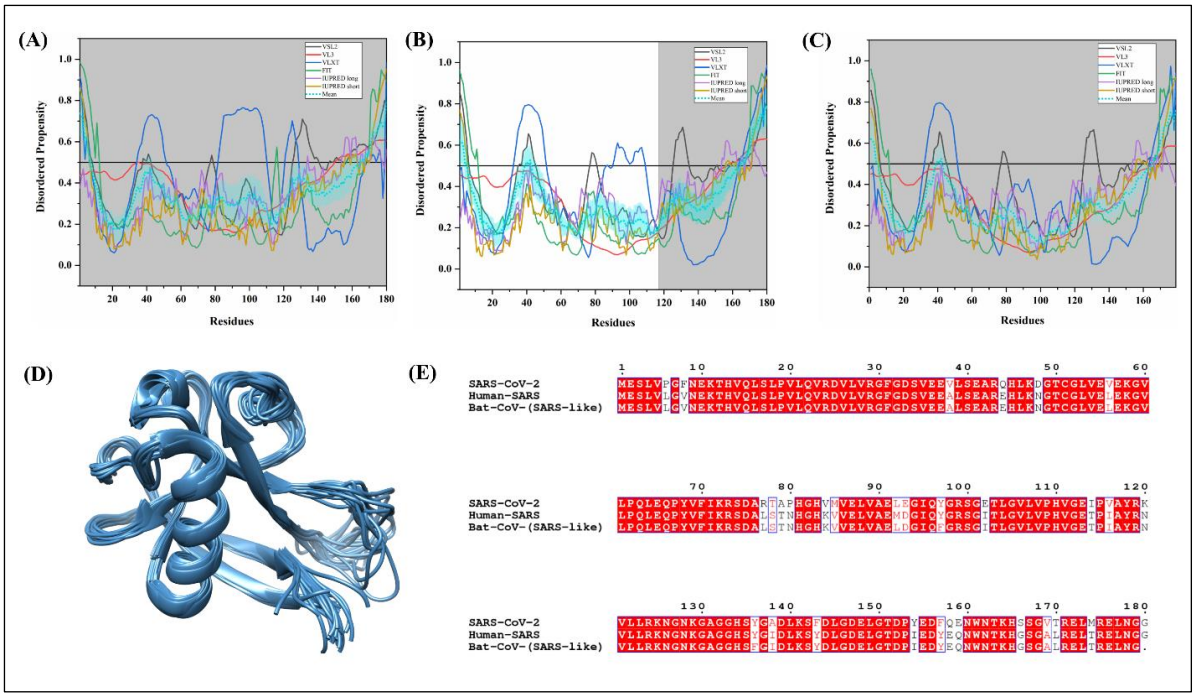

Figure 19. Analysis of intrinsic disorder propensity of Nsp1. (A) Graph represents the IDP analysis in SARS-CoV-2**, (B)** Human SARS, and **(C)** Bat CoV. The color schemes are same as figure 3.**(D)** Shows the PDB structure of NSP1 of Huma SARS generated by solution NMR (PDB ID: 2GDT) of 13-128 residues. **(E)** MSA profile of NSP1 of SARS-CoV-2, Human SARS, and Bat CoV. Red shaded sequences represent the similarity between the protein sequences.

SARS-CoV-2 nsp1 protein share 84.44% sequence identity with nsp1 of Human SARS and 83.80% with nsp1 of Bat CoV. Its N-terminal region is found to be quite conserved than the rest of the protein sequence (**Figure 19E)**. Mean PPIDs of nsp1 proteins of SARS-CoV-2, Human SARS, and Bat CoV are 12.78%, 14.44%, and 12.85%, respectively. **Figure 19A, 19B,** and **19C)** illustrate the graphs of predicted intrinsic disorder propensity in residues of Nsp1 proteins of SARS-CoV-2, Human SARS, and Bat CoV. According to the analysis, the following residues are predicted to have disorderedness, SARS-CoV-2 (1-7, 165-180), Human SARS (1-5, 165-180), and Bat CoV (1-5, 165-179). NMR-based structure of Nsp1 of Human SARS revealed the presence of two unstructured segments near the N-terminal (1-12 residues) and C-terminal (129-179 residues) regions (Almeida *et al*, 2007). The disordered region (128–180 residues) at C-terminus are characterized to be important for Nsp1 expression (Jauregui *et al*, 2013). Based on sequence homology with Human SARS Nsp1 protein, the predicted disordered Nsp1 C-terminal region of SARS-CoV-2 may play a critical role in its expression.

Non-structural protein 2 (Nsp2): This protein functions by disrupting the host survival pathway by interacting with host proteins Prohibitin1 and Prohibitin 2 (Cornillez-Ty *et al*, 2009). Reverse genetic deletion in the coding sequence of Nsp2 of the SARS virus attenuated little viral growth and replication and allowed the recovery of mutant virulent viruses. This indicates the dispensible nature of Nsp2protein in SARS viruses (Graham *et al*, 2005).

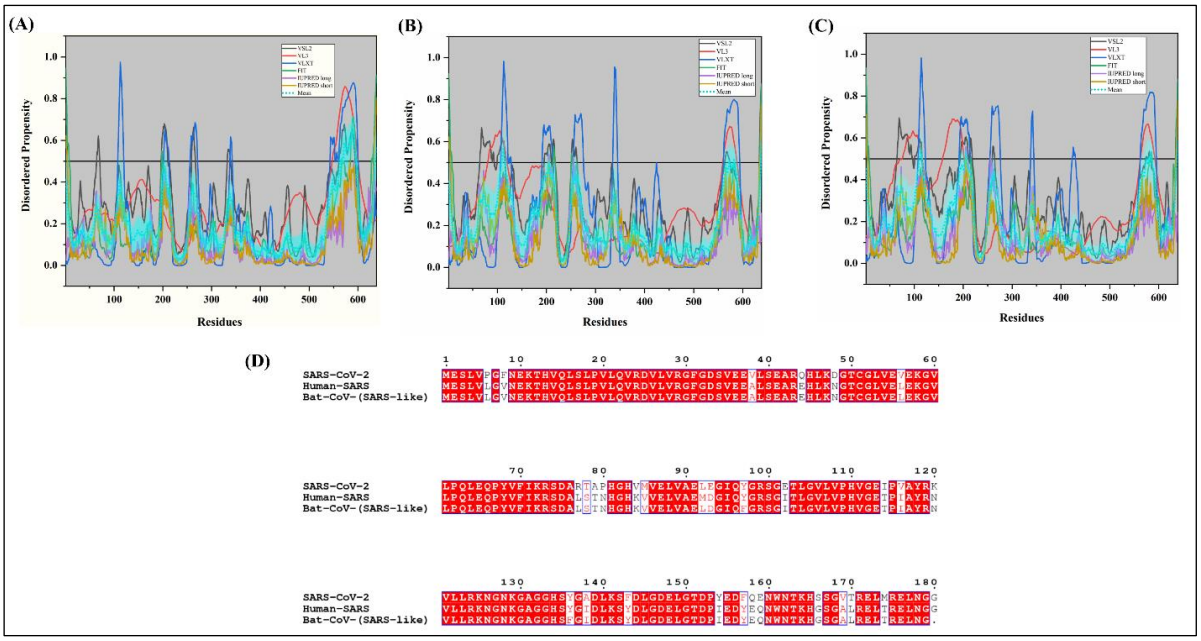

Figure 20. Analysis of intrinsic disorder propensity of Nsp2. (A) Graph represents the IDP analysis in SARS-CoV-2**, (B)** Human SARS, and **(C)** Bat CoV. The color schemes are same as figure 3. **(D)** MSA profile of Nsp2 of SARS-CoV-2, Human SARS, and Bat CoV. Red shaded sequences represent the similarity between the protein sequences.

Nsp2 protein is not found to be much conserved in SARS-CoV-2. However, it shares equal similarity with both the analogs having a percentage identity of 68.34% with Nsp2 of Human SARS and 68.97% with Nsp2 of Bat CoV (**Figure 20D**). We have estimated the mean PPIDs of nsp2 proteins of SARS-CoV-2, Human SARS, and Bat CoV to be 5.17%, 2.04%, and 2.03% respectively. The predisposition of intrinsic disorder in residues of nsp2 proteins of SARS-CoV-2, Human SARS, and Bat CoV are depicted in graphs in **figure 20A, 20B,** and **20C.** According to the analysis, the following residues in Nsp2 proteins are predicted to have disorderedness, SARS-CoV-2 (570-595), Human SARS (110-115), and Bat CoV (112-116).

Non-structural protein 3 (Nsp3): Nsp3 is a viral papain-like protease that affects the phosphorylation and activation of IRF3 and therefore antagonizes the IFN pathway. In another biochemical study, it was demonstrated that Nsp3 works by stabilizing NF-$\kappa\beta$ inhibitor further blocking the NF-$\kappa\beta$ pathway (Frieman *et al*, 2009). **Figure 21D** depicts the 1.85Å X-ray diffraction-based crystal structure (PDB ID: 2FE8) of the catalytic core of Nsp3 protein of Human SARS is obtained by Andrew and colleagues. This structure is consisting of 723-1036 residues. The structure revealed folds similar to a deubiquitinating enzyme in-vitro deubiquitinating activity of which was found to be efficiently high (Ratia *et al*, 2006).

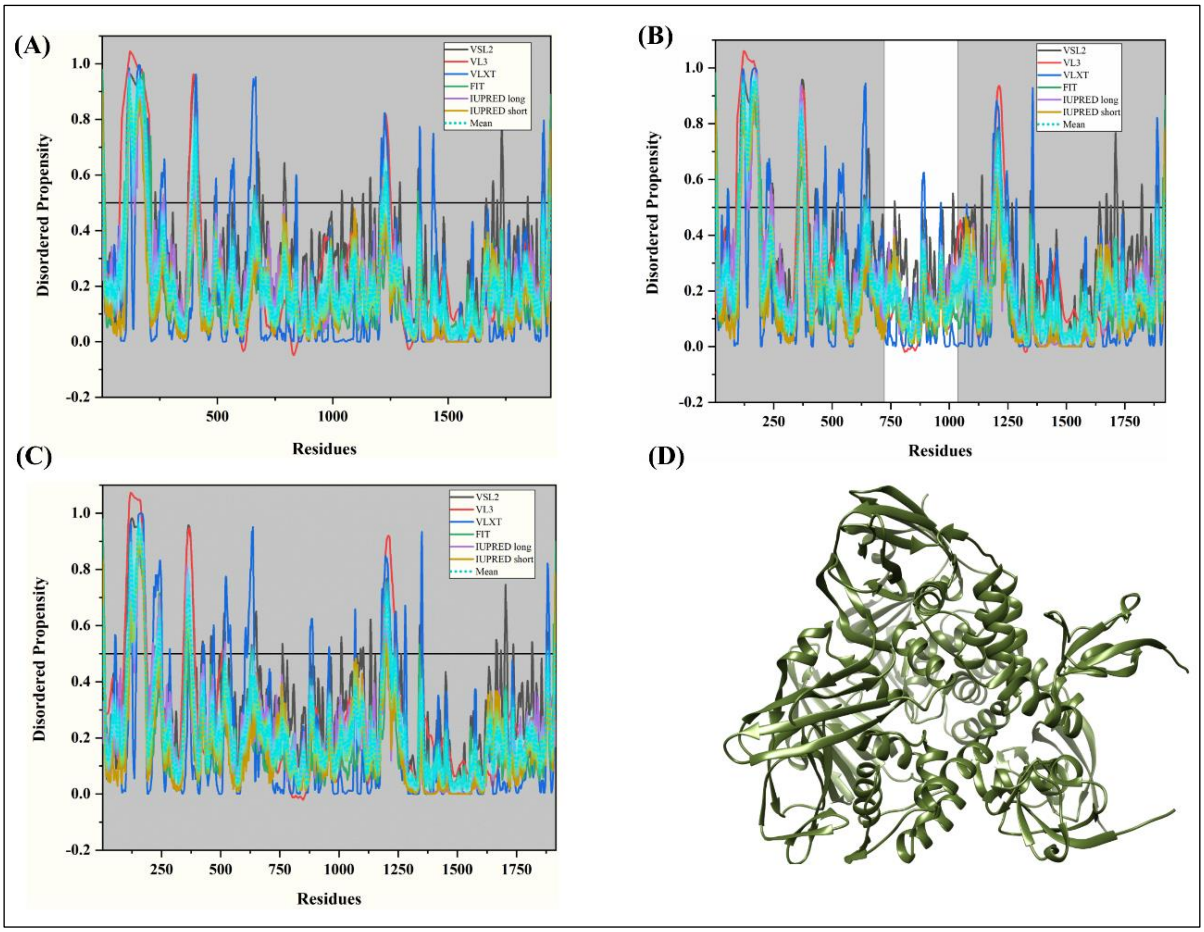

Figure 21. Analysis of intrinsic disorder propensity of Nsp3. (A) Graph represents the IDP analysis in SARS-CoV-2**, (B)** Human SARS and **(C)** Bat CoV. The color schemes are same as figure 3. **(D)** A 1.85 Å PDB structure of Nsp3 of Human SARS generated by X-ray diffraction of residues 723-1036 (dark olive colour) (PDB ID: 2FE8).

Nsp3 protein of SARS-CoV-2 contains several substituted residues throughout the protein. It is equally close with both Nsp3 proteins of Human SARS and Bat CoV sharing respective 76.69% and 76.31% identity **(Supplementary Figure S2A)**. According to our results, the mean PPIDs of Nsp3 proteins of SARS-CoV-2, Human SARS, and Bat CoV are 7.40%, 7.91%, and 7.78% respectively (**Table 2**). Graphs in **figure 21A, 21B,** and **21C** portray the tendency of intrinsic disorder in residues of Nsp3 proteins of SARS-CoV-2, Human SARS, and Bat CoV. Nsp3 proteins of all three studied SARS viruses were found to be highly structured. According to the analysis, following residues in Nsp3 proteins are predicted to have disorderedness, SARS-CoV-2 (1-5, 105-199, 1221-1238), Human SARS (102-189, 355-384, 1195-1223) and Bat CoV (107-182, 352-376, 1191-1217).

Non-structural protein 4 (Nsp4): Nsp4 has been reported to induce double-membrane vesicles (DMVs) with the co-expression of full-length Nsp3 and Nsp6 proteins for optimal replication inside host cells (Angelini *et al*, 2013; Hagemeijer *et al*, 2011; Sakai *et al*, 2017). It localizes itself in ER-membrane when expressed alone but is demonstrated to be present in replication units in infected cells. It was observed that Nsp4 protein forms a tetraspanning

transmembrane region having its N- and C-terminals in the cytosol (Oostra *et al*, 2007b). No crystal or NMR solution structure is reported for this protein.

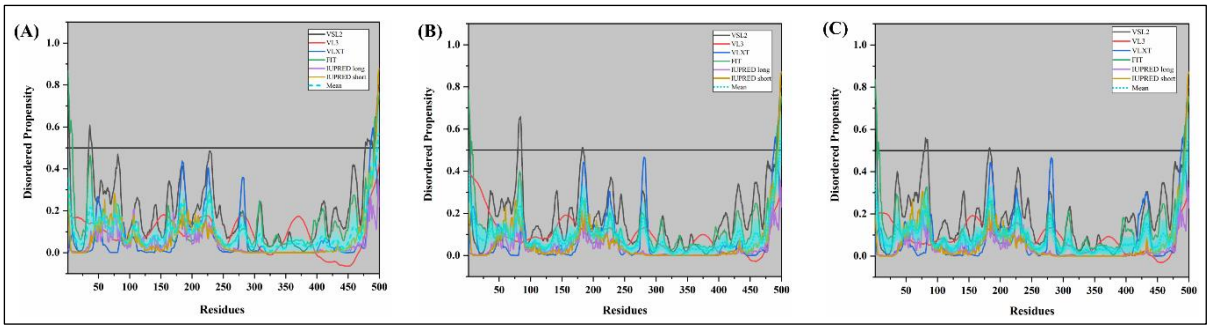

Figure 22. Analysis of intrinsic disorder propensity of Nsp4. (A) Graph represents the IDP analysis in SARS-CoV-2**, (B)** Human SARS, and **(C)** Bat CoV. The color schemes are same as figure 3.

Nsp4 protein of SARS-CoV-2 has multiple substitutions near the N-terminal region and has a quite conserved C-terminus **(Supplementary Figure S2B)**. It is found closer to Nsp4 of Bat CoV (81.40% identity) than Nsp4 of Human SARS (80%). Mean PPID of Nsp4proteins of SARS-CoV-2, Human SARS, and Bat CoV are estimated to be 0.80%, 0.60%, and 0.60% respectively. Predicted intrinsic disorder propensity in residues of Nsp4 proteins of SARS-CoV-2, Human SARS, and Bat CoV are depicted in graphs **figure 22A, 22B,** and **22C.** With PPIDs around zero, Nsp4 proteins were analyzed to be highly structured. According to the analysis, disorderedness is not found in the Nsp4 protein.

Non-structural protein 5 (Nsp5): Also referred to as 3CL-pro, Nsp5 works as a protease and cleaves replicase polyprotein (1a and 1ab) at 11 major sites (Tomar *et al*, 2015; Sparks *et al*, 2008). Crystal structure (PDB ID: 5C5O) obtained by X-ray diffraction from 3241-3546 residues is shown in **figure 23D.** In this complex, 3CL-protease is bind to a phenyl-beta-alanyl (S, R)-N-declin type inhibitor. Another crystal structure resolved to 1.96Å revealed a chymotrypsin-like fold and a conserved substrate-binding site connected to a novel α-helical fold (Anand *et al*, 2002). Recently, the structure of Nsp5 of SARS-CoV-2 has been crystallized using an X-ray diffraction-based technique (PDB ID: 6LU7) (**Figure 23E)**.

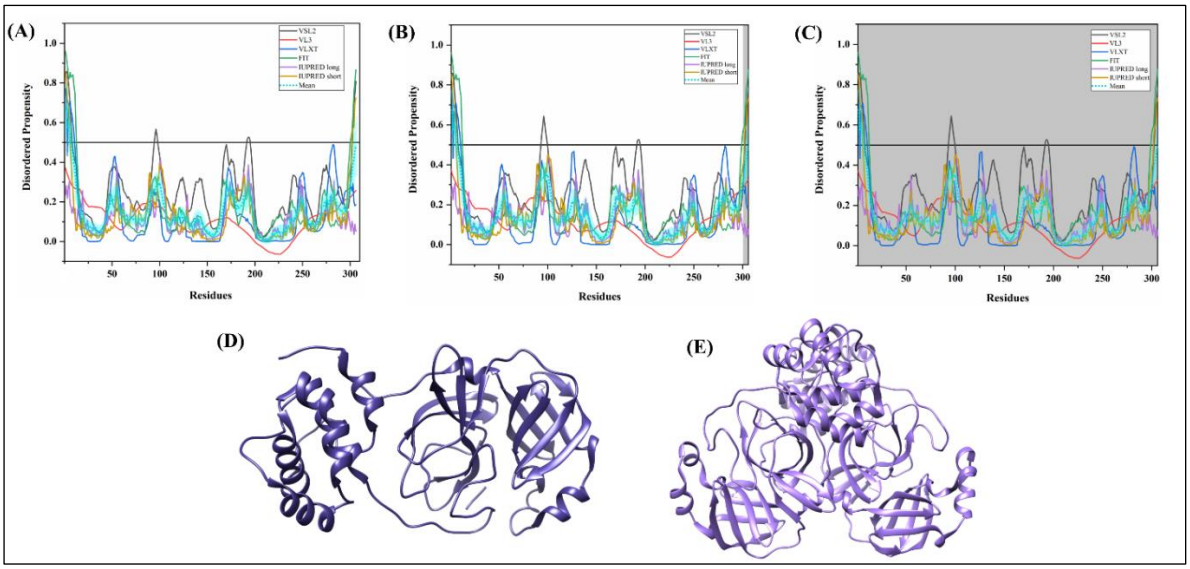

Figure 23. Analysis of intrinsic disorder propensity of Nsp5. (A) Graph represents the IDP analysis in SARS-CoV-2**, (B)** Human SARS, and **(C)** Bat CoV. The color schemes are same as figure 3. **(D)** A 1.50Å PDB structure of Nsp5 of Human SARS generated by X-ray diffraction (purple colour) (PDB ID: 5C5O) of 1- 300 residues. **(E)** Shows the PDB structure of NSP5 in SARS-CoV-2 generated by X-ray diffraction (purple colour) (PDB ID: 6LU7) of 1-306 residues.

Nsp5 protein is found conserved in all three studied SARS viruses. SARS-CoV-2 Nsp5 shares its 96.08% sequence identity with Nsp5 of Human SARS and 95.42% with Nsp5 of Bat CoV **(Supplementary Figure S2C)**.
Our results demonstrate the mean PPID of SARS-CoV-2, Human SARS, and Bat CoV to be equal to 1.96% (**Table 2**). The calculated predicted intrinsic disorder propensity in residues of respective (SARS-CoV-2, Human SARS, and Bat CoV Nsp5 proteins are illustrated in graphs of **figure 23A, 23B,** and **23C.** As the graphs depict, Nsp5 proteins were found to have no intrinsic disorder. According to the analysis, residues 1-6 have disorderedness in all three CoV.

Non-structural protein 6 (Nsp6): Nsp6 protein is involved in blocking ER-induced autophagosome/autolysosome vesicle that functions in restricting viral production inside host cells. It induces autophagy by activating the omegasome pathway which is normally implied by cells in response to starvation. SARS Nsp6 leads to the generation of small autophagosome vesicles thereby limiting their expansion (Cottam *et al*, 2014).

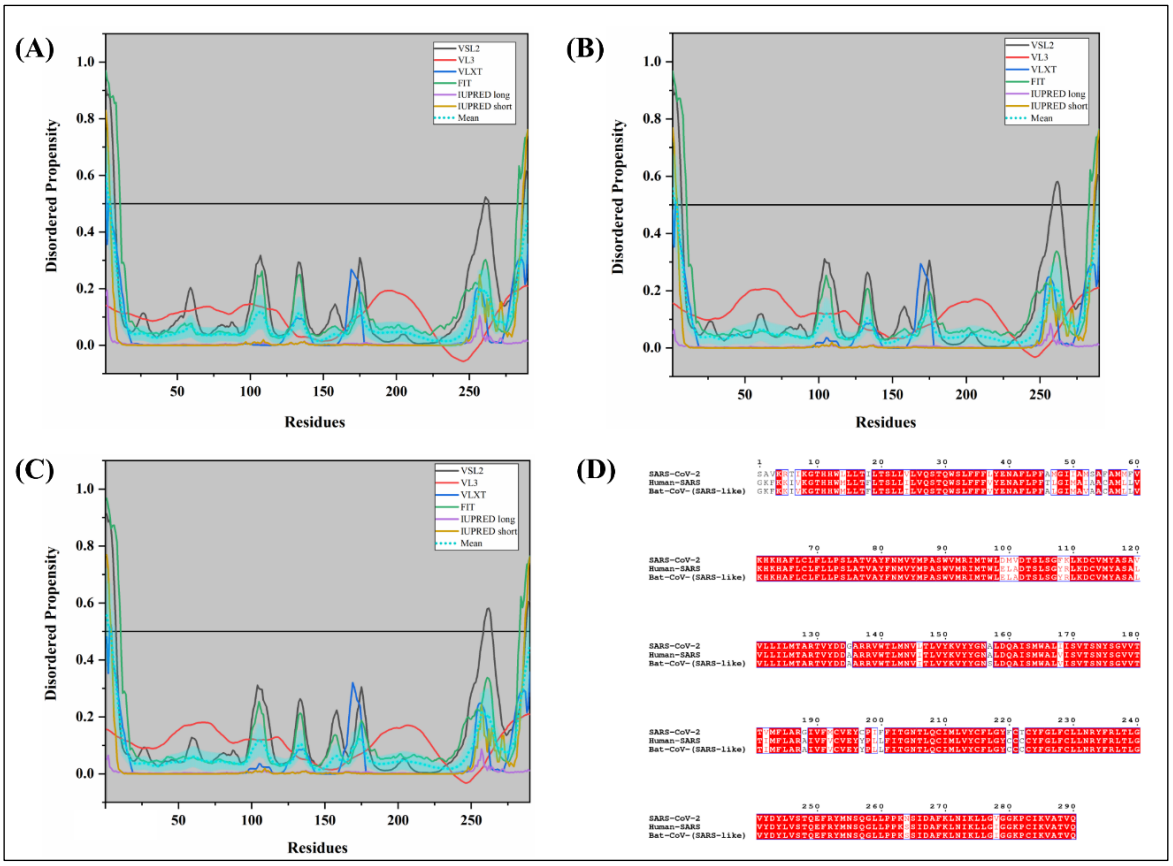

Figure 24. Analysis of intrinsic disorder propensity of Nsp6. (A) Graph represents the IDP analysis in SARS-CoV-2**, (B)** Human SARS and **(C)** Bat CoV. The color schemes are same as figure 3. **(D)** MSA profile of Nsp6 of SARS-CoV-2, Human SARS, and Bat CoV. Red shaded sequences represent the similarity between the protein sequences.

Nsp6 of SARS-CoV-2 is equally close to both Human SARS, and Bat CoV having a sequence identity of 87.24% (**Figure 24D**). According to our results, mean PPIDs in Nsp6 proteins are calculated to be 1.03% - SARS-CoV-2, 1.03% - Human SARS, and 4.48% - Bat CoV. **Figure 24A, 24B,** and **24C** show the respective graphs of intrinsic disorder tendency in Nsp6 proteins of SARS-CoV-2, Human SARS, and Bat CoV. From our predicted IDP results, we found no disorder in all three Nsp6 proteins. According to the analysis, disorderedness is not found in the Nsp4 protein.

Non-structural protein 7 and 8 (Nsp7 and 8): The ~10kDa Nsp7 protein helps in primase-independent *de novo* initiation of viral RNA replication by forming a hexadecameric ring-like structure with Nsp8 protein (te Velthuis *et al*, 2012; Zhai *et al*, 2005). Both non-structural proteins (7 and 8) contribute 8 molecules to the ring structured multimeric viral RNA polymerase. Site-directed mutagenesis in Nsp8 protein revealed a D/ExD/E motif essential for *in-vitro* catalysis (te Velthuis *et al*, 2012). **Figure 25D** depicts the 3.1Å resolution electron microscopy-based structure (PDB ID: 6NUR) of RDRP-nsp8-nsp7 complex bound to Nsp12 protein. The structure identified a conserved neutral Nsp7 and 8 binding sites overlapping with finger and thumb domains on Nsp12 of virus (Kirchdoerfer & Ward, 2019).

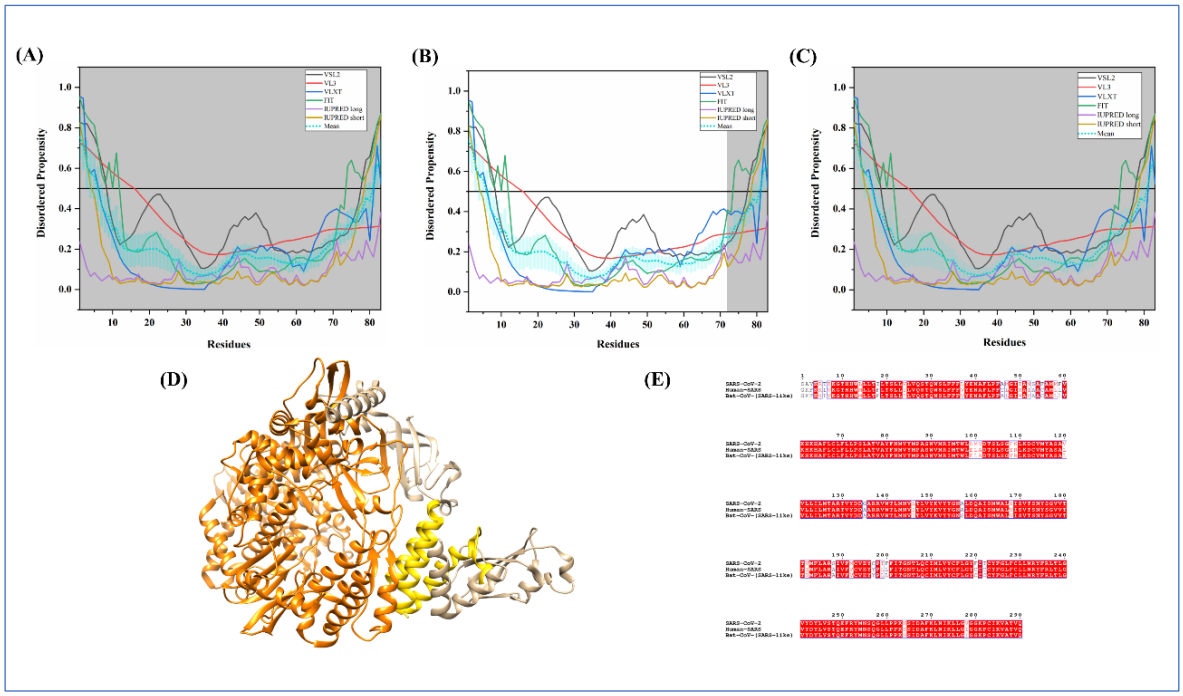

Figure 25. Analysis of intrinsic disorder propensity of Nsp7. (A) Graph represents the IDP analysis in SARS-CoV-2**, (B)** Human SARS, and **(C)** Bat CoV. The color schemes are same as figure 3. **(D)** A 3.10 Å resolution PDB structure of Nsp12-Nsp8-Nsp7 complex (PDB ID: 6NUR) generated by cryo-EM. Chain C represents nsp7 (gold colour) of 2-71 residues, B and D chains represent Nsp8 of 77-191 residues and chain A signifies Nsp12 (RNA-directed RNA polymerase) (orange colour) of 117-896 and 907-920 residues, of Human SARS. **(E)** MSA profile of NSP7 of SARS-CoV-2, Human SARS and Bat CoV. Red shaded sequences represent the similarity between the protein sequences.

We found that Nsp7 of SARS-CoV-2 share 100% sequence identity with Nsp7 of Bat CoV and 98.80% with Nsp7 Human SARS (**Figure 25E**) while novel Nsp8 protein is more closer to Nsp8 of Human SARS (97.47%) than Nsp8 of Bat CoV (96.46%) (**Figure 26D**).

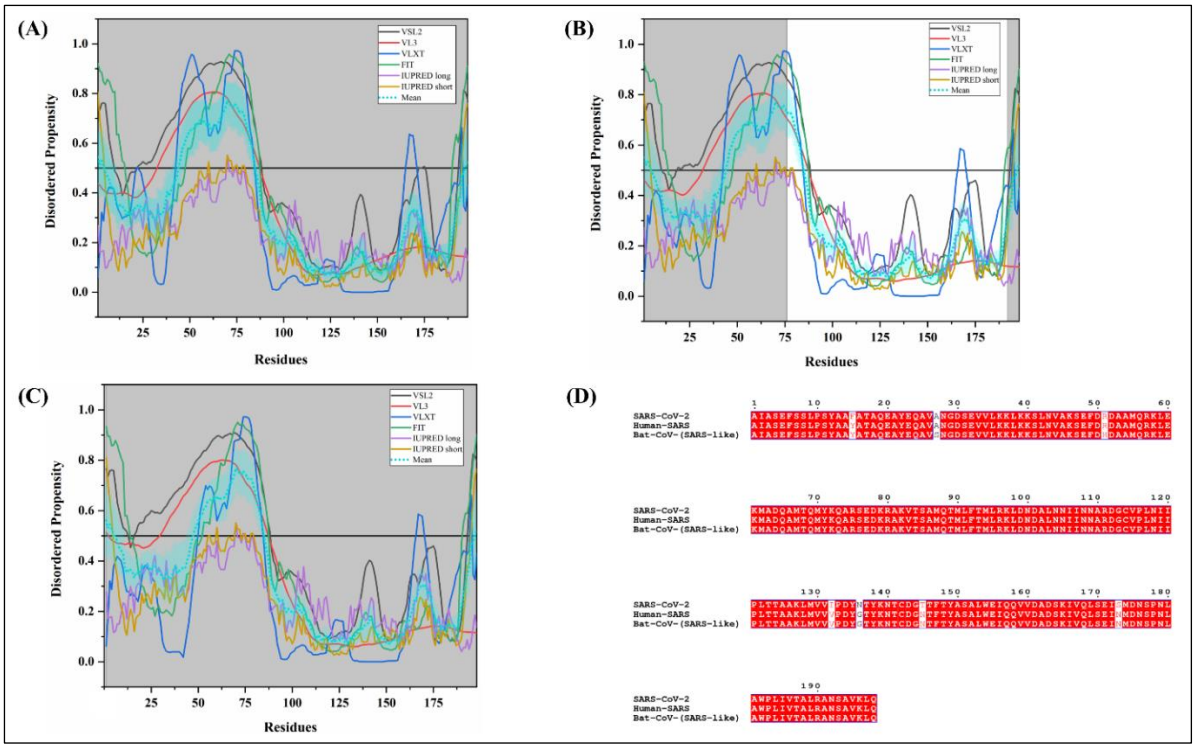

Figure 26. Analysis of intrinsic disorder propensity of Nsp8. (A) Graph represents the IDP analysis in SARS-CoV-2**, (B)** Human SARS and **(C)** Bat CoV. The color schemes are same as figure 3. **(D)** MSA profile of Nsp8 of SARS-CoV-2, Human SARS and Bat CoV. Red shaded sequences represent the similarity between the protein sequences.

Due to conserved residues, mean PPIDs of all Nsp7 proteins were found to be equal to 9.64%. Both SARS-CoV-2 and Human SARS Nsp8 proteins were calculated to have a mean PPID of 23.74% and, for Nsp8 of Bat CoV mean disorder is predicted to be 22.22%. **Figure 25A, 25B,** and **25C** display the graphs of predicted intrinsic disorder tendency in Nsp7 proteins of SARS-CoV-2, Human SARS, and Bat CoV. Graphs in **figure 26A, 26B,** and **26C** represent the predicted intrinsic disorder propensity in Nsp8 proteins of SARS-CoV-2, Human SARS, and Bat CoV. As our analysis suggests, Nsp7 proteins have a well-predicted structure while Nsp8 proteins have a moderate amount of disorderness. Nsp8 proteins is predicted to have a long disorder segment from 44-84 residues in both SARS-CoV-2 and Human SARS and 48-84 residues in Bat CoV.

Non-structural protein 9 (Nsp9): Nsp9 protein is a single-stranded RNA-binding protein (Egloff *et al*, 2004). It might provide protection from nucleases by binding and stabilizing viral nucleic acids during replication or transcription (Egloff *et al*, 2004). Presumed to evolve from a protease, Nsp9 forms a dimer using its GXXXG motif (Ponnusamy *et al*, 2008; Miknis *et al*, 2009). **Figure 27D** shows a 2.7Å crystal structure (PDB ID: 1QZ8) of Human SARS Nsp9 that identified an oligosaccharide/oligonucleotide fold-like fold in its structure (Egloff *et al*, 2004). Each monomer contains a cone-shaped β-barrel and a C-terminal α-helix arranged into a compact domain (Egloff *et al*, 2004).

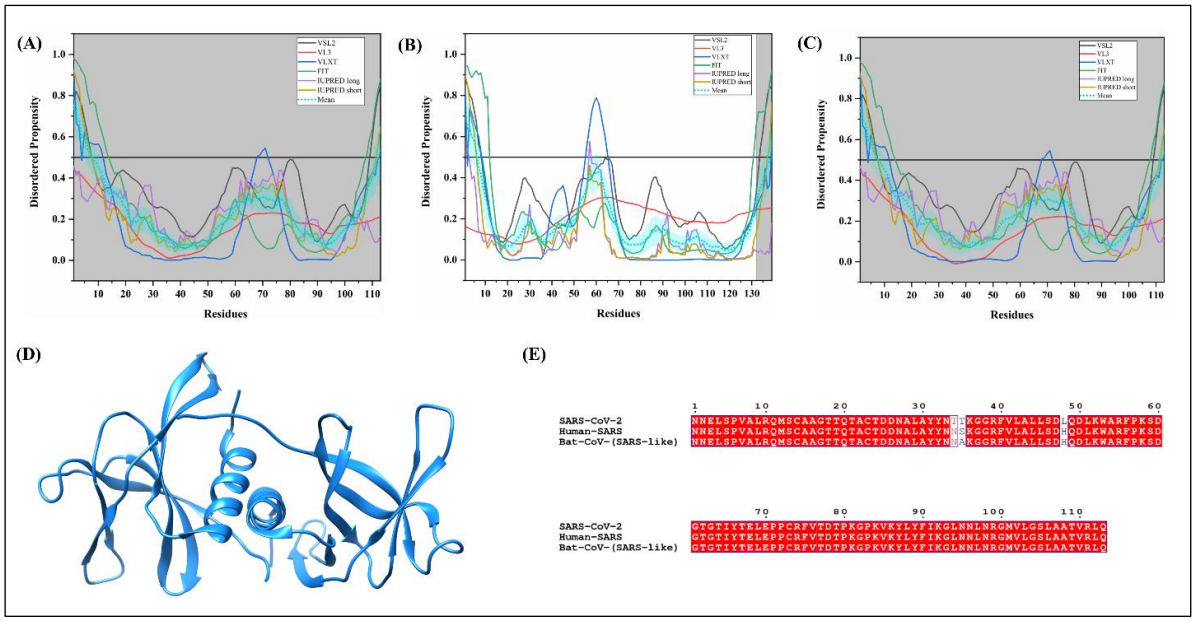

Figure 27. Analysis of intrinsic disorder propensity of Nsp9. (A) Graph represents the IDP analysis in SARS-CoV-2**, (B)** Human SARS and **(C)** Bat CoV. The color schemes are same as figure 3. **(D)** A 2.70 Å PDB structure of Nsp9 (dodger blue colour) (PDB ID: 1QZ8) of 3-113 residues of Human SARS generated by X-ray diffraction. **(E)** MSA profile of Nsp9 of SARS-CoV-2, Human SARS and Bat CoV. Red shaded sequences represent the similarity between the protein sequences.

Nsp9 of SARS-CoV-2 is equally identical to nsp9 proteins of both Human SARS and Bat CoV having a percentage identity of 97.35%. The difference in three amino acids at 34, 35 and 48 positions accounts for the above similarity score (**Figure 27E**). As calculated, the mean PPIDs of Nsp9 proteins of SARS-CoV-2, Human SARS, and Bat CoV are 7.08%, 7.96%, and 7.08% respectively. Graphs in **figures 27A, 27B,** and **27C** depict the predicted intrinsic disorder propensity in the Nsp9 protein of SARS-CoV-2, Human SARS, and Bat CoV. According to our analysis of intrinsic disorder, all three Nsp9 proteins are completely structured.

Non-structural protein 10 (Nsp10): Nsp10 performs several functions for SARS-CoV. It forms a complex with Nsp14 for hydrolyzes of dsRNA in 3′ to 5′ direction and activates in its exonuclease activity (Bouvet *et al*, 2012). It also stimulates the MTase activity of Nsp14 protein required during RNA-cap formation after replication (Bouvet *et al*, 2010). **Figure 28D** represents the structure (PDB ID: 5C8T) of the Nsp10/Nsp14 complex. In consistence with the previous biochemical experimental results, the structure identified important interactions with the ExoN (exonuclease domain) of Nsp14 without affecting its N7-Mtase activity (Bouvet *et al*, 2012, 2010).

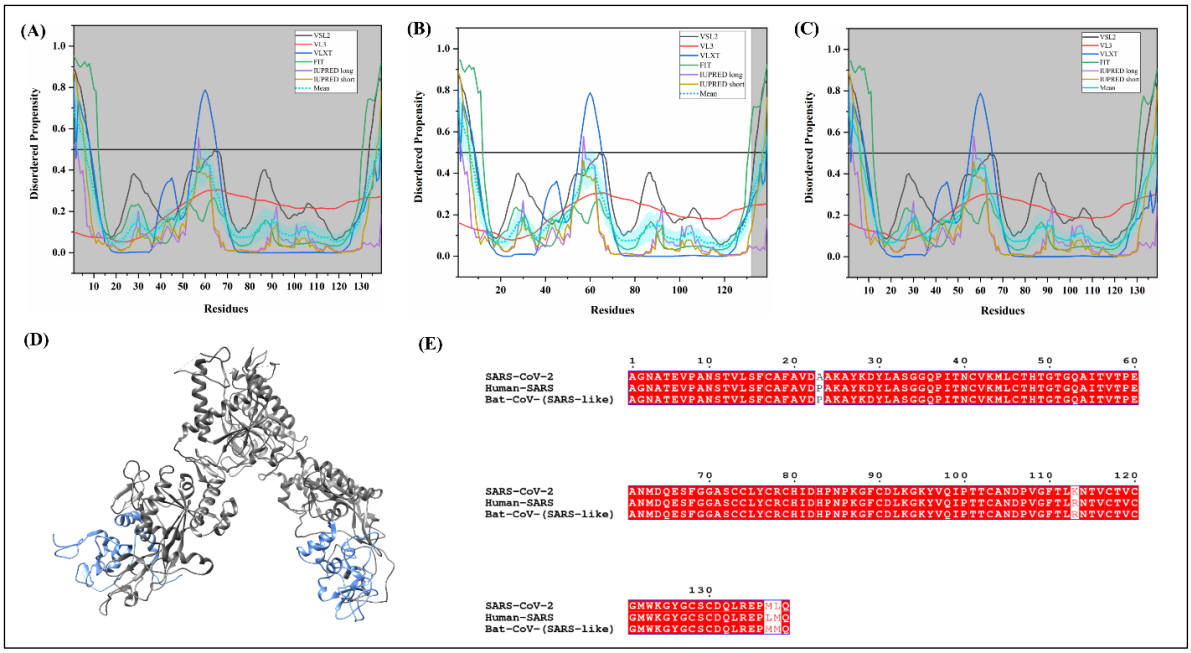

Figure 28. Analysis of intrinsic disorder propensity of Nsp10. (A) Graph represents the IDP analysis in SARS-CoV-2**, (B)** Human SARS, and **(C)** Bat CoV. The color schemes are same as figure 3. **(D)** A 3.20 Å PDB structure of Nsp10/Nsp14 complex (PDB ID: 5C8T) of Huma SARS generated by X-ray diffraction. In this structure, A and C chain (cornflower blue colour) signify Nsp10 of 1-131 residues and B and D chains are Nsp14 (dim grey colour) of 1-453 and 465-525 residues. **(E)** MSA profile of Nsp10 of SARS-CoV-2, Human SARS and Bat CoV. Red shaded sequences represent the similarity between the protein sequences.

SARS-CoV-2 Nsp10 protein is quite conserved having a 97.12% sequence identity with Nsp10 of Human SARS and 97.84% with Nsp10 of Bat CoV (**Figure 28E**). Mean PPIDs of all three studied Nsp10 proteins is found to be 5.04%. **Figures 28A, 28B,** and **28C** show the graphs of predicted intrinsic disorder tendency in Nsp10 proteins of SARS-CoV-2, Human SARS, and Bat CoV respectively. Our results of the predicted intrinsic disorder suggest the absence of disorder among studied Nsp10 proteins.

Non-structural protein 12 (Nsp12): Nsp12 is an RNA-dependent RNA Polymerase (RDRP) of coronaviruses. It carries out both primer-independent and primer-dependent synthesis of viral RNA with Mn^{2+} as its metallic co-factor and viral Nsp7 and 8 as protein co-factors (Ahn *et al*, 2012). As mentioned above, a 3.1Å resolution crystal structure (PDB ID: 6NUR) of Human SARS Nsp12 in association with Nsp7 and Nsp8 proteins has been reported using electron microscopy (**Figure 25D).** Nsp12 has a polymerase domain similar to "right hand", finger domain (398–581, 628–687 residues), palm domain (582–627, 688–815 residues) and a thumb domain (816–919) (Kirchdoerfer & Ward, 2019).

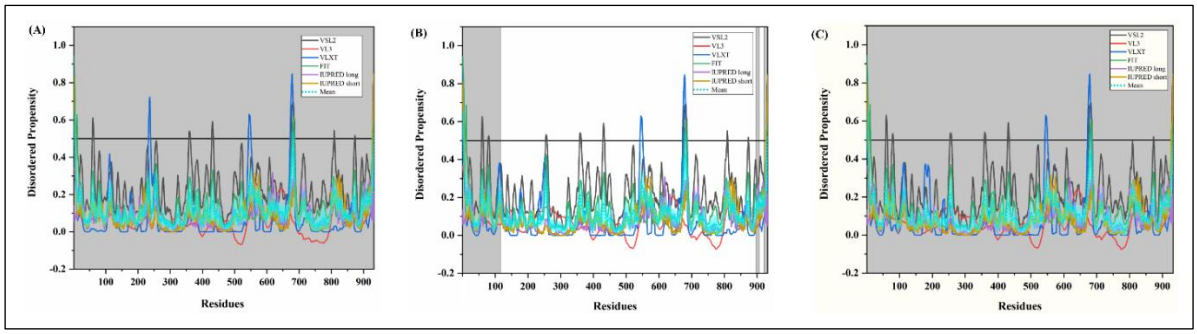

Figure 29. Analysis of intrinsic disorder propensity of Nsp12. (A) Graph represents the IDP analysis in SARS-CoV-2**, (B)** Human SARS, and **(C)** Bat CoV. The color schemes are same as figure 3.

SARS-CoV-2 Nsp12 protein has a conserved C-terminus **(Supplementary Figure S2D)**. It is found to share 96.35% sequence identity with Human SARS Nsp12 protein and 95.60% with Bat CoV Nsp12. Mean PPID in all three Nsp12 proteins is estimated to be 0.43% (**Table 2**). Graphs in **figure 29A, 29B,** and **29C** illustrate the respective predicted intrinsic disorder in Nsp12 proteins of SARS-CoV-2, Human SARS, and Bat CoV. As expected, no significant disorder was predicted in Nsp12 proteins.

Non-structural protein 13 (Nsp13): Nsp13 functions as a viral helicase and unwinds dsDNA/dsRNA in 5' to 3' polarity (Adedeji *et al*, 2012). Recombinant Helicase expressed in E.coli Rosetta 2 strain has been reported to unwind ~280 bp per second (Adedeji *et al*, 2012). **Figure 30D** represents a 2.8Å X-ray diffraction-based crystal structure (PDB ID: 6JYT) of Human SARS Nsp13 protein. Helicase contains a β19-β20 loop on 1A domain which is primarily responsible for its unwinding activity. Moreover, the study revealed an important interaction of Nsp12 protein with Nsp13 that further enhances its helicase activity (Jia *et al*, 2019).

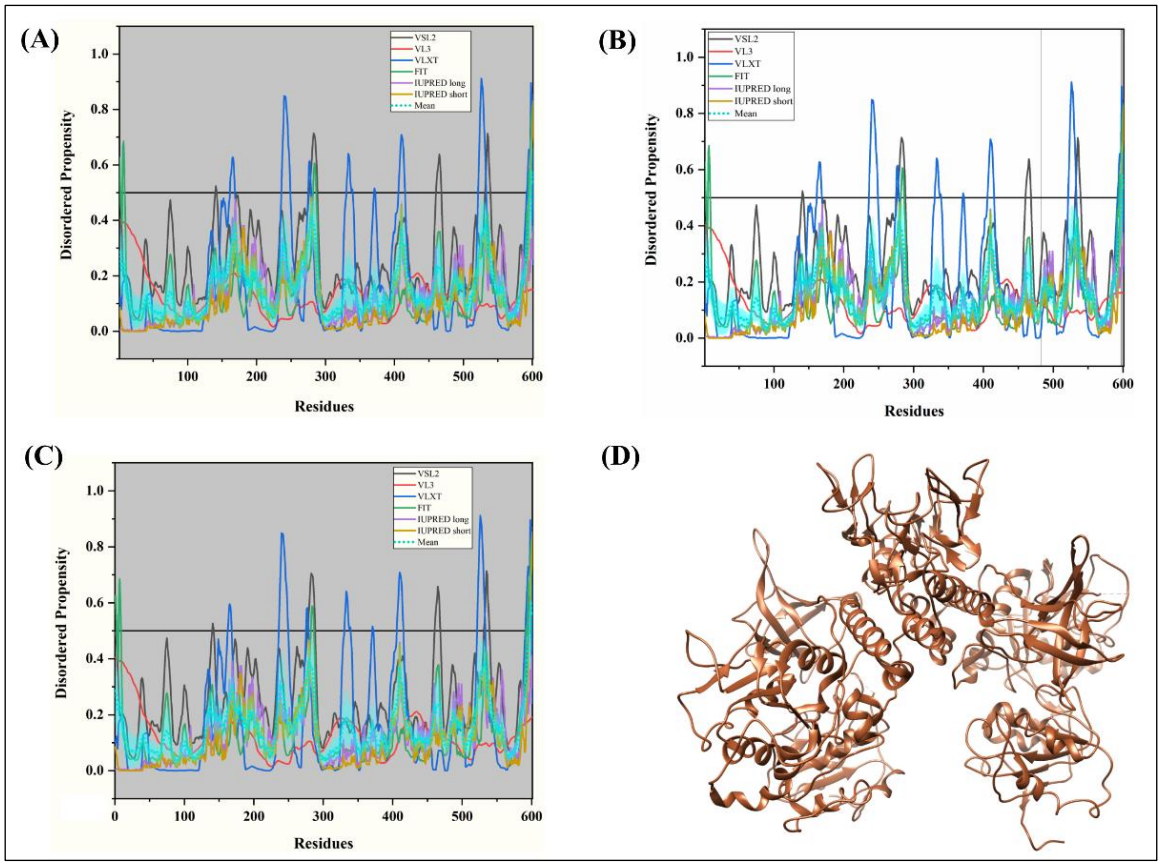

Figure 30. Analysis of intrinsic disorder propensity of Nsp13. (A) Graph represents the IDP analysis in SARS-CoV-2**, (B)** Human SARS, and **(C)** Bat CoV. The color schemes are same as figure 3. **(D)** A 2.80 Å PDB structure of Nsp13 of Huma SARS (PDB ID: 6JYT) (sienna colour) of 1-596 residues, generated by X-ray diffraction.

The 601 amino acid long Nsp13 protein of SARS-CoV-2 is nearly conserved as it shares 99.83% with Nsp13 of Humans SARS and 98.84% with Nsp13 of Bat CoV **(Supplementary Figure S2E)**. In accordance with our results, the mean PPIDs of all three Nsp13 proteins are estimated to be 0.67%. **Figure 30A, 30B,** and **30C** show the respective graphs for the predisposition of intrinsic disorder in Nsp13 proteins of SARS-CoV-2, Human SARS, and Bat CoV. **Figure 30D** shows PDB structure of helicase/Nsp13 (PDB ID: 6JYT), from 5302-5902 residues (sienna colour). No significant disorder is revealed in all three Nsp13 proteins.

Non-structural protein 14: Multifunctional nsp14 acts as an Exoribonuclease (ExoN) and methyltransferase (N7-Mtase) for SARS-Coronaviruses. It's 3’ to 5’ exonuclease activity lies in conserved DEDD residues related to exonuclease superfamily (Minskaia *et al*, 2006). Its guanine-N7 methyltransferase activity depends upon the S-adenosyl-L-methionine (AdoMet) as a cofactor (Bouvet *et al*, 2010). As mentioned previously, nsp14 requires nsp10 protein in for activating its ExoN and N7-Mtase activity inside host cells. **Figure 28D** depicts the 3.2Å deduced crystal structure (PDB ID: 5C8T) of human SARS nsp10/nsp14 complex using the X-ray diffraction technique. Amino acids 1-287 form the ExoN domain and 288-527 residues from the N7-Mtase domain of nsp14. A loop present from 288-301 residues is essential for its N7-Mtase activity (Ma *et al*, 2015).

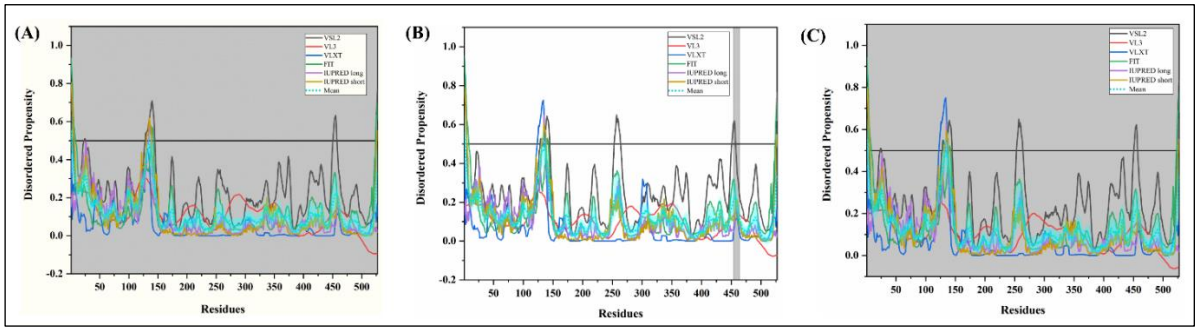

Figure 31. Analysis of intrinsic disorder propensity of NSP14. (A) Graph represents the IDP analysis in SARS-CoV-2**, (B)** Human SARS and **(C)** Bat CoV (SARS-like). The color schemes are same as figure 3.

SARS-CoV-2 nsp14 protein shares a 95.07% percentage identity with Human SARS nsp14 and 94.69% with Bat CoV nsp14 (**Supplementary Figure S2F**). Mean PPID of nsp14 proteins of SARS-CoV-2 and Human SARS is calculated to be 0.38% while nsp14 of Bat CoV has a mean PPID 0.57%. Predicted intrinsic disorder propensity in residues of nsp14 proteins of SARS-CoV-2, Human SARS, and Bat CoV are represented in graphs in **figure 31A, 10B,** and **31C**. As can be observed from PPID values, all nsp14 proteins are found to be highly structured.

Non-structural protein 15 (Nsp15): Nsp15 is a uridylate-specific RNA Endonuclease (NendoU) that creates a 2′–3′ cyclic phosphates after cleavage. Its endonuclease activity depends upon Mn^{2+} ions as co-factors. Conserved in Nidovirus, it acts as an important genetic marker due to its absence in other RNA viruses (Ivanov *et al*, 2004). As illustrated in **figure 32D**, Bruno and colleagues deduced a 2.6Å crystal structure (PDB ID: 2H85) of Uridylate-specific Nsp15 using X-ray diffraction. The monomeric Nsp15 has three domains: N-terminal domain (1-62 residues) formed by a three anti-parallel β-sheets and two α-helices packed together; a middle domain (63–191 residues) contains an α-helix connect to a 39 amino acids long coil to two α and five β region; and a C-terminal domain (192–345 residues) consisting of two anti-parallel three β-sheets on each side of a central α-helical core (Ricagno *et al*, 2006).

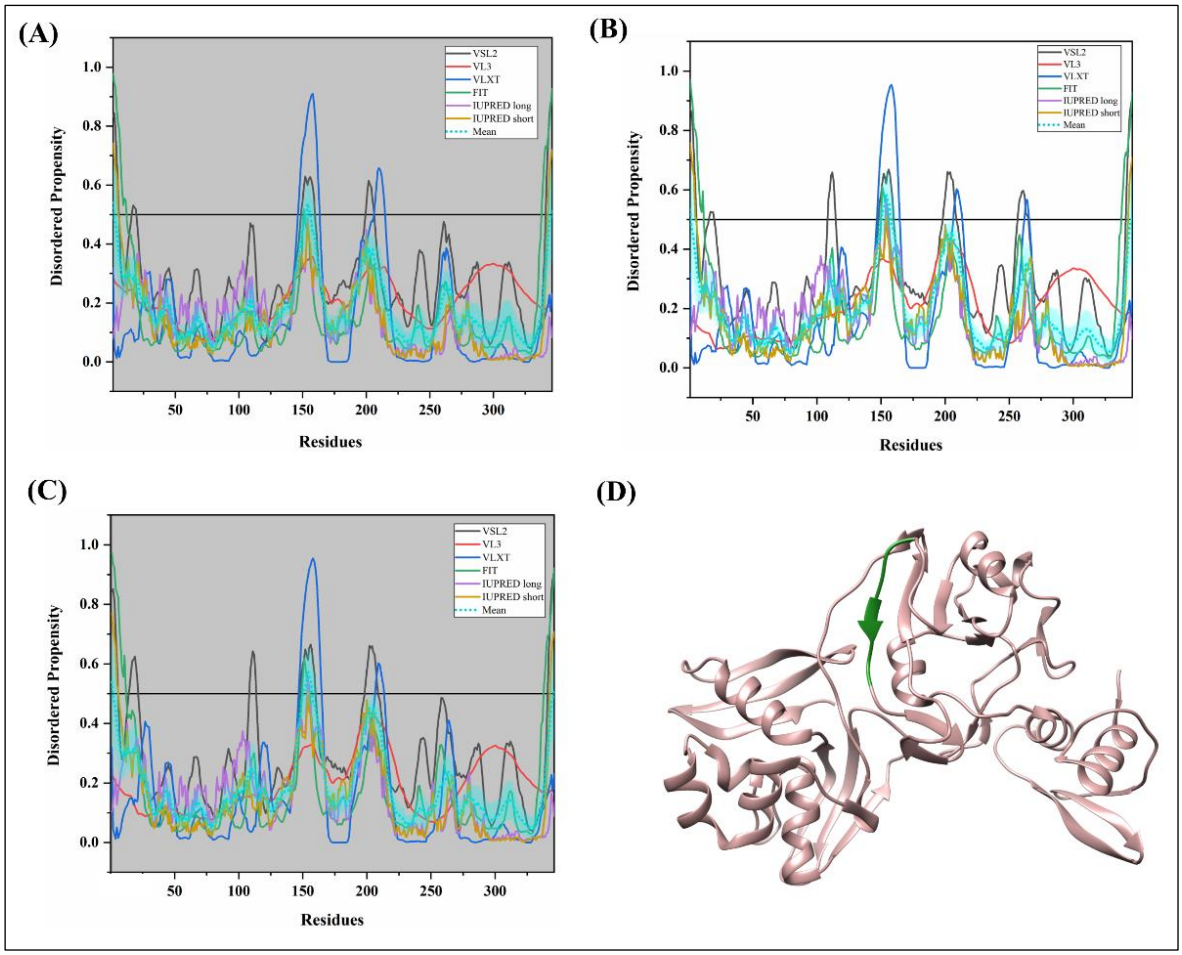

Figure 31. Analysis of intrinsic disorder propensity of Nsp15. (A) Graph represents the IDP analysis in SARS-CoV-2**, (B)** Human SARS, and **(C)** Bat CoV. The color schemes are same as figure 3. **(D)** A 2.60 Å PDB structure generated by X-ray diffraction of Nsp15 (PDB ID:2H85) (rosy brown colour) of 1-346 residues. Residues of 151-157 have the disorder in their structure represented in forest-green colour.

Nsp15 protein is found to be quite conserved across human and bat CoVs. It shares an 88.73% sequence identity with Nsp15 of Human SARS and 88.15% with Nsp15 of Bat CoV (**Supplementary Figure S2G**). Calculated mean PPID of SARS-CoV-2, Human SARS, and Bat CoV is 1.73%, 2.60%, and 2.60%, respectively. Predicted intrinsic disorder propensity in residues of Nsp15 proteins of SARS-CoV-2 Human SARS and Bat CoV are represented in graphs in **figure 32A, 32B,** and **32C.** No significant disorder was revealed in studied Nsp15 proteins.

Non-structural protein 16 (Nsp16): Nsp16 protein is another Mtase domain-containing protein. As methylation of coronavirus mRNAs occurs in steps, three protein Nsp10, Nsp14 and Nsp16 acts one after another. The first event requires the initiation trigger from Nsp10 protein after which Nsp14 methylates capped mRNAs forming cap-0 (7Me) GpppA-RNAs. Nsp16 protein along with its co-activator protein Nsp10 acts on cap-0 (7Me) GpppA-RNAs to give rise to final cap-1 (7Me)GpppA(2'OMe)-RNAs (Bouvet *et al*, 2010; Decroly *et al*, 2008). A 2Å X-ray diffraction-based structure (PDB ID: 3R24) of the Human SARS nsp10-

nsp16 complex is depicted in **figure 33D.** The structure consists of a characteristic fold present in class I MTase family comprising of α-helices and loops surrounding a seven-stranded β-sheet (Chen *et al*, 2011).

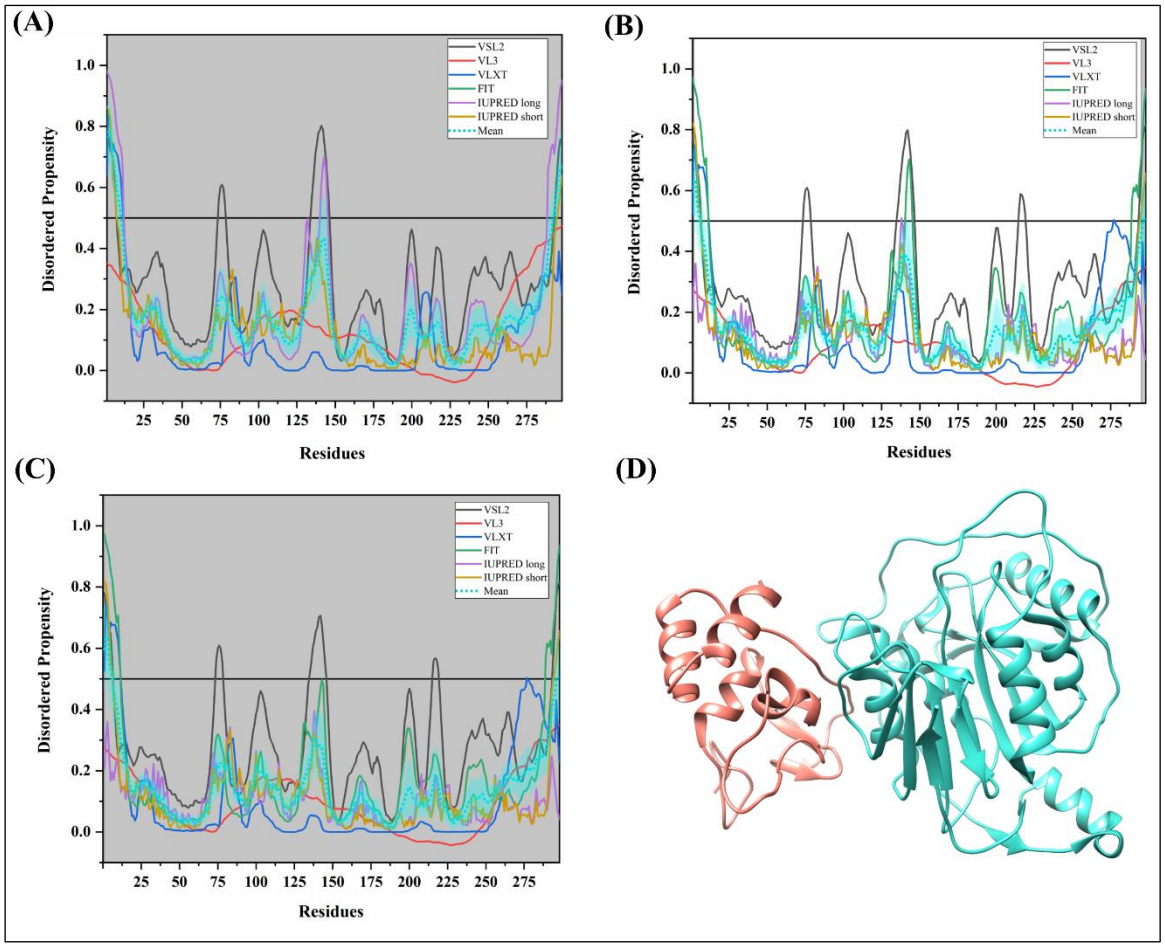

Figure 33. Analysis of intrinsic disorder propensity of Nsp16. (A) Graph represents the IDP analysis in SARS-CoV-2**, (B)** Human SARS and **(C)** Bat CoV. The color schemes are same as figure 3. **(D)** A 2.60 Å the PDB structure of the nsp10-nsp16 complex (PDB ID: 3R24) of Human SARS generated by X-ray diffraction. Chain A signifies Nsp16 (turquoise colour) of 3-294 residues.

Nsp16 protein of SARS-CoV-2 is found equally similar with Nsp16 proteins of Human SARS and Bat CoV (93.29%) (**Supplementary Figure S2H**). As observed using different predictors, mean PPIDs SARS-CoV-2, Human SARS, and Bat CoV are 5.37%, 3.02%, and 3.02%. Graphs in **figure 33A, 33B,** and **33C** show the predicted predisposition for intrinsic disorder in nsp16 proteins of SARS-CoV-2, Human SARS and Bat CoV, respectively. We found no disorder in Nsp16 proteins which are observed to be structured.

Replicase polyprotein 1a: As 1a contains identical non-structural proteins 1-10 with 1ab, therefore, we have not performed their IDP analysis separately. However, it has one additional non-structural protein designated as Nsp11.

Non-structural protein 11 (Nsp11): This is an uncharacterized protein cleaved from replicase polyprotein 1a. The small protein with unknown function requires experimental insights to further characterize this protein. The software used in this study requires >30 amino acid sequence, therefore, due to short sequence of all three studied coronaviruses Nsp11 proteins do not show any disordered residue**.** MSA of Nsp11 protein of SARS-CoV-2 is found to have a similarity of 84.62% with Nsp11 proteins of Human SARS and Bat CoV (**Figure 34**).

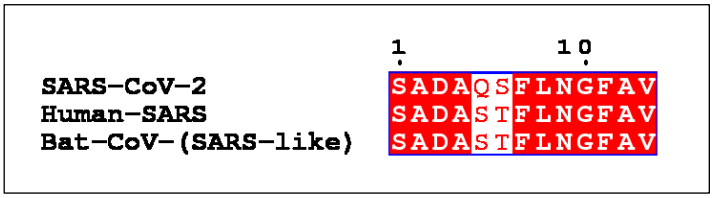

Figure 34. MSA profile of Nsp11 of SARS-CoV-2, Human SARS and Bat CoV. Red shaded sequences represent the similarity between the protein sequences.

Concluding remarks:

The emergence of viruses and associated deaths around the globe is a major concern to mankind. There is very little information available in the public domain regarding protein structure and functions of SARS-CoV-2 yet. Based on the similarity with Human SARS and Bat CoV, the published reports have suggested the functions of its proteins. Using available information on its genome and translated proteome from Genbank, we have carried out a comprehensive analysis of disorderedness present in proteins of SARS-CoV-2. Additionally, a comparison is also made with its close relatives from the same group of beta coronaviruses; Human SARS and Bat CoV. As result suggested, the N proteins are predicted to be highly disordered having more than 60% of PPID. Another moderately disordered protein is encoded from ORF6, which downregulates the interferon pathway. All other proteins have shown less disordered regions depicting a three-dimensional structure in the native state. Generally, IDPs undergo structural transition upon association with their physiological partners, therefore, this study will help to understand the interaction of other viral proteins as well as host proteins in different physiological conditions. This will also guide structural biologists to carry out a structure-based analysis of its genome to explore the path for the development of new drugs and vaccines.

Future perspective:

The periodical outbreaks of pathogens worldwide always remind the lack of suitable drugs or vaccines for proper cure or treatment. In 2003, nearly 750 deaths were reported due to the SARS outbreak in more than 24 countries, but this time, the outbreak of Wuhan's novel coronavirus has surpassed this number quickly indicating more causalities in near future. The lack of accurate information and ignorance of primary symptoms are major reasons which cause many infection cases. The actual reason for SARS-CoV-2 spread is still unknown but some assumptions made by researchers and Chinese authorities, and also confirmed its transmission from human to human. It also has made major impacts on education and

economy worldwide due to several restrictions such as traveling, transportation. Due to advancements in distinguished techniques, the full genome sequence was made available in a few days of the first infection report from Wuhan, China. But further research needs to be done to identify its actual cause and suitable treatment in the coming future. There are certain possibilities that can be explored with the available information. More in-depth experimental studies using molecular and cell biology to establish structure-function relationships are required to understand its proper functioning. Additionally, based on homology and other information on protein-protein interaction, the associated viral and host proteins should be explored which help in carrying out replication, maturation, and ultimately pathogenesis. Using structural biology, various purposes including drug development could be achieved such as high throughput screening of compounds virtually as well as experimentally. The thorough experimental disorder analysis of three coronaviruses in this study will also help to structure biologists to design experiments keeping this information in mind.

Authors Contribution: RG: Conception and design, interpretation of data, writing, and review of the manuscript and study supervision. MS, TB, PK, BRG, KG: acquisition and interpretation of data, writing of the manuscript.

Acknowledgments: All the authors would like to thank IIT Mandi for providing facilities. MS and BRG were supported by MHRD for funding. KG was supported by the Department of Biotechnology (DBT), India (BT/PR16871/NER/95/329/2015). PK was supported by IIT Mandi-IIT Ropar-PGI Chandigarh, BioX consortium grant (IITM/INT/RG/18). TB is grateful to the Department of Science and Technology for her INSPIRE fellowship for Funding.

Conflict of Interest: All authors declare that there is no financial competing interest.

References:

Adedeji AO, Marchand B, Te Velthuis AJW, Snijder EJ, Weiss S, Eoff RL, Singh K & Sarafianos SG (2012) Mechanism of nucleic acid unwinding by SARS-CoV helicase. *PLoS One* **7:** e36521 Available at: http://www.ncbi.nlm.nih.gov/pubmed/22615777 [Accessed February 29, 2020]

Ahn D-G, Choi J-K, Taylor DR & Oh J-W (2012) Biochemical characterization of a recombinant SARS coronavirus nsp12 RNA-dependent RNA polymerase capable of copying viral RNA templates. *Arch. Virol.* **157:** 2095–104 Available at: http://www.ncbi.nlm.nih.gov/pubmed/22791111 [Accessed February 29, 2020]

Almeida MS, Johnson MA, Herrmann T, Geralt M & Wuthrich K (2007) Novel -Barrel Fold in the Nuclear Magnetic Resonance Structure of the Replicase Nonstructural Protein 1 from the Severe Acute Respiratory Syndrome Coronavirus. *J. Virol.* **81:** 3151–3161

Anand K, Palm GJ, Mesters JR, Siddell SG, Ziebuhr J & Hilgenfeld R (2002) Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. *EMBO J.* **21:** 3213–24 Available at: http://www.ncbi.nlm.nih.gov/pubmed/12093723 [Accessed February 29, 2020]

Angelini MM, Akhlaghpour M, Neuman BW & Buchmeier MJ (2013) Severe acute respiratory syndrome coronavirus nonstructural proteins 3, 4, and 6 induce double-membrane vesicles. *MBio* **4:** Available at:

http://www.ncbi.nlm.nih.gov/pubmed/23943763 [Accessed February 29, 2020]

Belouzard S, Millet JK, Licitra BN & Whittaker GR (2012) Mechanisms of coronavirus cell entry mediated by the viral spike protein. *Viruses* **4:** 1011–1033

Bouvet M, Debarnot C, Imbert I, Selisko B, Snijder EJ, Canard B & Decroly E (2010) In vitro reconstitution of SARS-coronavirus mRNA cap methylation. *PLoS Pathog.* **6:** e1000863 Available at: http://www.ncbi.nlm.nih.gov/pubmed/20421945 [Accessed February 29, 2020]

Bouvet M, Imbert I, Subissi L, Gluais L, Canard B & Decroly E (2012) RNA 3'-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. *Proc. Natl. Acad. Sci. U. S. A.* **109:** 9372–7 Available at: http://www.ncbi.nlm.nih.gov/pubmed/22635272 [Accessed February 29, 2020]

Broer R, Boson B, Spaan W, Cosset F-L & Corver J (2006) Important Role for the Transmembrane Domain of Severe Acute Respiratory Syndrome Coronavirus Spike Protein during Entry. *J. Virol.* **80:** 1302–1310

Cavanagh D & Davis PJ (1986) Coronavirus IBV: removal of spike glycopolypeptide S1 by urea abolishes infectivity and haemagglutination but not attachment to cells. *J. Gen. Virol.* **67 (Pt 7):** 1443–8 Available at: http://www.ncbi.nlm.nih.gov/pubmed/3014054 [Accessed February 29, 2020]

Chang C-K, Hsu Y-L, Chang Y-H, Chao F-A, Wu M-C, Huang Y-S, Hu C-K & Huang T-H (2009) Multiple Nucleic Acid Binding Sites and Intrinsic Disorder of Severe Acute Respiratory Syndrome Coronavirus Nucleocapsid Protein: Implications for Ribonucleocapsid Protein Packaging. *J. Virol.* **83:** 2255–2264

Chen Y, Su C, Ke M, Jin X, Xu L, Zhang Z, Wu A, Sun Y, Yang Z, Tien P, Ahola T, Liang Y, Liu X & Guo D (2011) Biochemical and structural insights into the mechanisms of SARS coronavirus RNA ribose 2'-O-methylation by nsp16/nsp10 protein complex. *PLoS Pathog.* **7:** e1002294 Available at: http://www.ncbi.nlm.nih.gov/pubmed/22022266 [Accessed February 29, 2020]

Chinese SARS Molecular Epidemiology Consortium (2004) Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. *Science* **303:** 1666–9 Available at: http://www.ncbi.nlm.nih.gov/pubmed/14752165 [Accessed March 7, 2020]

Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J & Sayers EW (2016) GenBank. *Nucleic Acids Res.* **44:** D67–D72

Corman VM, Lienau J & Witzenrath M (2019) [Coronaviruses as the cause of respiratory infections]. *Internist (Berl).* **60:** 1136–1145 Available at: http://www.ncbi.nlm.nih.gov/pubmed/31455974 [Accessed February 29, 2020]

Cornillez-Ty CT, Liao L, Yates JR, Kuhn P & Buchmeier MJ (2009) Severe Acute Respiratory Syndrome Coronavirus Nonstructural Protein 2 Interacts with a Host Protein Complex Involved in Mitochondrial Biogenesis and Intracellular Signaling. *J. Virol.* **83:** 10314–10318

Coronavirus disease 2019 Available at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [Accessed February 29, 2020]

Corse E & Machamer CE (2003) The cytoplasmic tails of infectious bronchitis virus E and M proteins mediate their interaction. *Virology* **312:** 25–34

Cottam EM, Whelband MC & Wileman T (2014) Coronavirus NSP6 restricts autophagosome expansion. *Autophagy* **10:** 1426–41 Available at: http://www.ncbi.nlm.nih.gov/pubmed/24991833 [Accessed February 29, 2020]

Cotten M, Lam TT, Watson SJ, Palser AL, Petrova V, Grant P, Pybus OG, Rambaut A, Guan Y, Pillay D, Kellam P & Nastouli E (2013) Full-genome deep sequencing and

phylogenetic analysis of novel human betacoronavirus. *Emerg. Infect. Dis.* **19:** 736–742

Decroly E, Imbert I, Coutard B, Bouvet M, Selisko B, Alvarez K, Gorbalenya AE, Snijder EJ & Canard B (2008) Coronavirus Nonstructural Protein 16 Is a Cap-0 Binding Enzyme Possessing (Nucleoside-2'O)-Methyltransferase Activity. *J. Virol.* **82:** 8071–8084

DeDiego ML, Alvarez E, Almazan F, Rejas MT, Lamirande E, Roberts A, Shieh W-J, Zaki SR, Subbarao K & Enjuanes L (2007) A Severe Acute Respiratory Syndrome Coronavirus That Lacks the E Gene Is Attenuated In Vitro and In Vivo. *J. Virol.* **81:** 1701–1713

Dunker AK, Brown CJ & Obradovic Z (2002) Identification and functions of usefully disordered proteins. *Adv. Protein Chem.* **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. *FEBS J.* **272:** 5129–48 Available at: http://www.ncbi.nlm.nih.gov/pubmed/16218947 [Accessed February 29, 2020]

Dunker AK, Silman I, Uversky VN & Sussman JL (2008) Function and structure of inherently disordered proteins. *Curr. Opin. Struct. Biol.* **18:** 756–764

Egloff M-P, Ferron F, Campanacci V, Longhi S, Rancurel C, Dutartre H, Snijder EJ, Gorbalenya AE, Cambillau C & Canard B (2004) The severe acute respiratory syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world. *Proc. Natl. Acad. Sci. U. S. A.* **101:** 3792–6 Available at: http://www.ncbi.nlm.nih.gov/pubmed/15007178 [Accessed February 29, 2020]

Fan K, Wei P, Feng Q, Chen S, Huang C, Ma L, Lai B, Pei J, Liu Y, Chen J & Lai L (2004) Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase. *J. Biol. Chem.* **279:** 1637–42 Available at: http://www.ncbi.nlm.nih.gov/pubmed/14561748 [Accessed March 7, 2020]

Fielding BC, Tan Y-J, Shuo S, Tan THP, Ooi E-E, Lim SG, Hong W & Goh P-Y (2004) Characterization of a Unique Group-Specific Protein (U122) of the Severe Acute Respiratory Syndrome Coronavirus. *J. Virol.* **78:** 7311–7318

Frieman M, Ratia K, Johnston RE, Mesecar AD & Baric RS (2009) Severe Acute Respiratory Syndrome Coronavirus Papain-Like Protease Ubiquitin-Like Domain and Catalytic Domain Regulate Antagonism of IRF3 and NF- B Signaling. *J. Virol.* **83:** 6689–6705

Frieman M, Yount B, Heise M, Kopecky-Bromberg SA, Palese P & Baric RS (2007) Severe Acute Respiratory Syndrome Coronavirus ORF6 Antagonizes STAT1 Function by Sequestering Nuclear Import Factors on the Rough Endoplasmic Reticulum/Golgi Membrane. *J. Virol.* **81:** 9812–9824

Gadhave K, Gehi BR, Kumar P, Xue B, Uversky VN & Giri R (2020) The dark side of Alzheimer's disease: unstructured biology of proteins from the amyloid cascade signaling pathway. *Cell. Mol. Life Sci.* Available at: http://www.ncbi.nlm.nih.gov/pubmed/31894361 [Accessed February 29, 2020]

Garg N, Kumar P, Gadhave K & Giri R (2019) The dark proteome of cancer: Intrinsic disorderedness and functionality of HIF-1α along with its interacting proteins. *Prog. Mol. Biol. Transl. Sci.* **166:** 371–403 Available at: http://www.ncbi.nlm.nih.gov/pubmed/31521236 [Accessed February 29, 2020]

Giri R, Kumar D, Sharma N & Uversky VN (2016) Intrinsically Disordered Side of the Zika Virus Pro(1) Giri, R.; Kumar, D.; Sharma, N.; Uversky, V. N. Intrinsically Disordered Side of the Zika Virus Proteome. Frontiers in cellular and infection microbiology 2016, 6, 144.teome. *Front. Cell. Infect. Microbiol.* **6:** 144 Available at: http://www.ncbi.nlm.nih.gov/pubmed/27867910 [Accessed February 29, 2020]

Goh GK-M, Dunker AK & Uversky V (2013) Prediction of Intrinsic Disorder in MERS-CoV/HCoV-EMC Supports a High Oral-Fecal Transmission. *PLoS Curr.* **5:** Available

at: http://www.ncbi.nlm.nih.gov/pubmed/24270586 [Accessed February 29, 2020]

Gorbalenya AE, Enjuanes L, Ziebuhr J & Snijder EJ (2006) Nidovirales: evolving the largest RNA virus genome. *Virus Res.* **117:** 17–37 Available at: http://www.ncbi.nlm.nih.gov/pubmed/16503362 [Accessed February 29, 2020]

Graham RL & Baric RS (2010) Recombination, Reservoirs, and the Modular Spike: Mechanisms of Coronavirus Cross-Species Transmission. *J. Virol.* **84:** 3134–3146

Graham RL, Sims AC, Brockway SM, Baric RS & Denison MR (2005) The nsp2 Replicase Proteins of Murine Hepatitis Virus and Severe Acute Respiratory Syndrome Coronavirus Are Dispensable for Viral Replication. *J. Virol.* **79:** 13399–13411

Gunalan V, Mirazimi A & Tan Y-J (2011) A putative diacidic motif in the SARS-CoV ORF6 protein influences its subcellular localization and suppression of expression of co-transfected expression constructs. *BMC Res. Notes* **4:** 446 Available at: http://www.ncbi.nlm.nih.gov/pubmed/22026976 [Accessed February 29, 2020]

de Haan CAM, te Lintelo E, Li Z, Raaben M, Wurdinger T, Bosch BJ & Rottier PJM (2006) Cooperative Involvement of the S1 and S2 Subunits of the Murine Coronavirus Spike Protein in Receptor Binding and Extended Host Range. *J. Virol.* **80:** 10909–10918

Hagemeijer MC, Ulasli M, Vonk AM, Reggiori F, Rottier PJM & de Haan CAM (2011) Mobility and Interactions of Coronavirus Nonstructural Protein 4. *J. Virol.* **85:** 4572–4577

Huang C, Ito N, Tseng C-TK & Makino S (2006) Severe Acute Respiratory Syndrome Coronavirus 7a Accessory Protein Is a Viral Structural Protein. *J. Virol.* **80:** 7287–7294

Huang F, Oldfield C, Meng J, Hsu WL, Xue B, Uversky VN, Romero P & Dunker AK (2012) Subclassifying disordered proteins by the CH-CDF plot method. In *Pacific Symposium on Biocomputing* pp 128–139.

Huang Q, Yu L, Petros AM, Gunasekera A, Liu Z, Xu N, Hajduk P, Mack J, Fesik SW & Olejniczak ET (2004) Structure of the N-terminal RNA-binding domain of the SARS CoV nucleocapsid protein. *Biochemistry* **43:** 6059–63 Available at: http://www.ncbi.nlm.nih.gov/pubmed/15147189 [Accessed February 29, 2020]

Hussain S, Pan J, Chen Y, Yang Y, Xu J, Peng Y, Wu Y, Li Z, Zhu Y, Tien P & Guo D (2005) Identification of Novel Subgenomic RNAs and Noncanonical Transcription Initiation Signals of Severe Acute Respiratory Syndrome Coronavirus. *J. Virol.* **79:** 5288–5295

Ivanov KA, Hertzig T, Rozanov M, Bayer S, Thiel V, Gorbalenya AE & Ziebuhr J (2004) Major genetic marker of nidoviruses encodes a replicative endoribonuclease. *Proc. Natl. Acad. Sci. U. S. A.* **101:** 12694–12699

Jauregui AR, Savalia D, Lowry VK, Farrell CM & Wathelet MG (2013) Identification of residues of SARS-CoV nsp1 that differentially affect inhibition of gene expression and antiviral signaling. *PLoS One* **8:** e62416 Available at: http://www.ncbi.nlm.nih.gov/pubmed/23658627 [Accessed March 7, 2020]

Jia Z, Yan L, Ren Z, Wu L, Wang J, Guo J, Zheng L, Ming Z, Zhang L, Lou Z & Rao Z (2019) Delicate structural coordination of the Severe Acute Respiratory Syndrome coronavirus Nsp13 upon ATP hydrolysis. *Nucleic Acids Res.* **47:** 6538–6550 Available at: http://www.ncbi.nlm.nih.gov/pubmed/31131400 [Accessed February 29, 2020]

Kanzawa N, Nishigaki K, Hayashi T, Ishii Y, Furukawa S, Niiro A, Yasui F, Kohara M, Morita K, Matsushima K, Le MQ, Masuda T & Kannagi M (2006) Augmentation of chemokine production by severe acute respiratory syndrome coronavirus 3a/X1 and 7a/X4 proteins through NF-kappaB activation. *FEBS Lett.* **580:** 6807–12 Available at: http://www.ncbi.nlm.nih.gov/pubmed/17141229 [Accessed February 29, 2020]

Keng C-T, Choi Y-W, Welkers MRA, Chan DZL, Shen S, Gee Lim S, Hong W & Tan Y-J (2006) The human severe acute respiratory syndrome coronavirus (SARS-CoV) 8b

protein is distinct from its counterpart in animal SARS-CoV and down-regulates the expression of the envelope protein in infected cells. *Virology* **354:** 132–42 Available at: http://www.ncbi.nlm.nih.gov/pubmed/16876844 [Accessed February 29, 2020]

Khan S, Fielding BC, Tan THP, Chou C-F, Shen S, Lim SG, Hong W & Tan Y-J (2006) Over-expression of severe acute respiratory syndrome coronavirus 3b protein induces both apoptosis and necrosis in Vero E6 cells. *Virus Res.* **122:** 20–7 Available at: http://www.ncbi.nlm.nih.gov/pubmed/16965829 [Accessed February 29, 2020]

Kirchdoerfer RN & Ward AB (2019) Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat. Commun.* **10:** 2342 Available at: http://www.ncbi.nlm.nih.gov/pubmed/31138817 [Accessed February 29, 2020]

Kopecky-Bromberg SA, Martinez-Sobrido L, Frieman M, Baric RA & Palese P (2007) Severe Acute Respiratory Syndrome Coronavirus Open Reading Frame (ORF) 3b, ORF 6, and Nucleocapsid Proteins Function as Interferon Antagonists. *J. Virol.* **81:** 548–557

Kopecky-Bromberg SA, Martinez-Sobrido L & Palese P (2006) 7a Protein of Severe Acute Respiratory Syndrome Coronavirus Inhibits Cellular Protein Synthesis and Activates p38 Mitogen-Activated Protein Kinase. *J. Virol.* **80:** 785–793

Law HKW, Cheung CY, Ng HY, Sia SF, Chan YO, Luk W, Nicholls JM, Peiris JSM & Lau YL (2005a) Chemokine up-regulation in SARS-coronavirus-infected, monocyte-derived human dendritic cells. *Blood* **106:** 2366–74 Available at: http://www.ncbi.nlm.nih.gov/pubmed/15860669 [Accessed March 7, 2020]

Law PTW, Wong C-H, Au TCC, Chuck C-P, Kong S-K, Chan PKS, To K-F, Lo AWI, Chan JYW, Suen Y-K, Chan HYE, Fung K-P, Waye MMY, Sung JJY, Lo YMD & Tsui SKW (2005b) The 3a protein of severe acute respiratory syndrome-associated coronavirus induces apoptosis in Vero E6 cells. *J. Gen. Virol.* **86:** 1921–30 Available at: http://www.ncbi.nlm.nih.gov/pubmed/15958670 [Accessed February 29, 2020]

Van Der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, Fuxreiter M, Gough J, Gsponer J, Jones DT, Kim PM, Kriwacki RW, Oldfield CJ, Pappu R V., Tompa P, Uversky VN, Wright PE & Babu MM (2014) Classification of intrinsically disordered regions and proteins. *Chem. Rev.* **114:** 6589–6631 Available at: http://www.ncbi.nlm.nih.gov/pubmed/24773235 [Accessed February 29, 2020]

Li F, Li W, Farzan M & Harrison SC (2005) Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **309:** 1864–8 Available at: http://www.ncbi.nlm.nih.gov/pubmed/16166518 [Accessed February 29, 2020]

Li Y, Surya W, Claudine S & Torres J (2014) Structure of a conserved Golgi complex-targeting signal in coronavirus envelope proteins. *J. Biol. Chem.* **289:** 12535–49 Available at: http://www.ncbi.nlm.nih.gov/pubmed/24668816 [Accessed February 29, 2020]

Liu J, Perumal NB, Oldfield CJ, Su EW, Uversky VN & Dunker AK (2006) Intrinsic disorder in transcription factors. *Biochemistry* **45:** 6873–88 Available at: http://www.ncbi.nlm.nih.gov/pubmed/16734424 [Accessed February 29, 2020]

Liu J, Sun Y, Qi J, Chu F, Wu H, Gao F, Li T, Yan J & Gao GF (2010) The membrane protein of severe acute respiratory syndrome coronavirus acts as a dominant immunogen revealed by a clustering region of novel functionally and structurally defined cytotoxic T-lymphocyte epitopes. *J. Infect. Dis.* **202:** 1171–80 Available at: http://www.ncbi.nlm.nih.gov/pubmed/20831383 [Accessed February 29, 2020]

Lokugamage KG, Narayanan K, Huang C & Makino S (2012) Severe Acute Respiratory Syndrome Coronavirus Protein nsp1 Is a Novel Eukaryotic Translation Inhibitor That Represses Multiple Steps of Translation Initiation. *J. Virol.* **86:** 13598–13608

Lu W, Zheng B-J, Xu K, Schwarz W, Du L, Wong CKL, Chen J, Duan S, Deubel V & Sun B (2006) Severe acute respiratory syndrome-associated coronavirus 3a protein forms an

ion channel and modulates virus release. *Proc. Natl. Acad. Sci. U. S. A.* **103:** 12540–5 Available at: http://www.ncbi.nlm.nih.gov/pubmed/16894145 [Accessed February 29, 2020]

Ma Y, Wu L, Shaw N, Gao Y, Wang J, Sun Y, Lou Z, Yan L, Zhang R & Rao Z (2015) Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex. *Proc. Natl. Acad. Sci. U. S. A.* **112:** 9436–41 Available at: http://www.ncbi.nlm.nih.gov/pubmed/26159422 [Accessed February 29, 2020]

Masters PS (2006) The molecular biology of coronaviruses. *Adv. Virus Res.* **66:** 193–292 Available at: http://www.ncbi.nlm.nih.gov/pubmed/16877062 [Accessed February 29, 2020]

McBride CE, Li J & Machamer CE (2007) The Cytoplasmic Tail of the Severe Acute Respiratory Syndrome Coronavirus Spike Protein Contains a Novel Endoplasmic Reticulum Retrieval Signal That Binds COPI and Promotes Interaction with Membrane Protein. *J. Virol.* **81:** 2418–2428

McBride R & Fielding BC (2012) The role of severe acute respiratory syndrome (SARS)-coronavirus accessory proteins in virus pathogenesis. *Viruses* **4:** 2902–23 Available at: http://www.ncbi.nlm.nih.gov/pubmed/23202509 [Accessed February 29, 2020]

McBride R, van Zyl M & Fielding BC (2014) The coronavirus nucleocapsid is a multifunctional protein. *Viruses* **6:** 2991–3018

Meier C, Aricescu AR, Assenberg R, Aplin RT, Gilbert RJC, Grimes JM & Stuart DI (2006) The crystal structure of ORF-9b, a lipid binding protein from the SARS coronavirus. *Structure* **14:** 1157–65 Available at: http://www.ncbi.nlm.nih.gov/pubmed/16843897 [Accessed February 29, 2020]

Mészáros B, Erdos G & Dosztányi Z (2018) IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **46:** W329–W337 Available at: http://www.ncbi.nlm.nih.gov/pubmed/29860432 [Accessed February 29, 2020]

Miknis ZJ, Donaldson EF, Umland TC, Rimmer RA, Baric RS & Schultz LW (2009) Severe Acute Respiratory Syndrome Coronavirus nsp9 Dimerization Is Essential for Efficient Viral Growth. *J. Virol.* **83:** 3007–3018

Minskaia E, Hertzig T, Gorbalenya AE, Campanacci V, Cambillau C, Canard B & Ziebuhr J (2006) Discovery of an RNA virus 3'->5' exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **103:** 5108–13 Available at: http://www.ncbi.nlm.nih.gov/pubmed/16549795 [Accessed February 29, 2020]

Nal B, Chan C, Kien F, Siu L, Tse J, Chu K, Kam J, Staropoli I, Crescenzo-Chaigne B, Escriou N, van der Werf S, Yuen K-Y & Altmeyer R (2005) Differential maturation and subcellular localization of severe acute respiratory syndrome coronavirus surface proteins S, M and E. *J. Gen. Virol.* **86:** 1423–34 Available at: http://www.ncbi.nlm.nih.gov/pubmed/15831954 [Accessed February 29, 2020]

Narayanan K, Chen C-J, Maeda J & Makino S (2003) Nucleocapsid-independent specific viral RNA packaging via viral envelope protein and viral RNA signal. *J. Virol.* **77:** 2922–7 Available at: http://www.ncbi.nlm.nih.gov/pubmed/12584316 [Accessed February 29, 2020]

Narayanan K, Huang C & Makino S (2008) SARS coronavirus accessory proteins. *Virus Res.* **133:** 113–21 Available at: http://www.ncbi.nlm.nih.gov/pubmed/18045721 [Accessed March 7, 2020]

Nelson CA, Pekosz A, Lee CA, Diamond MS & Fremont DH (2005) Structure and intracellular targeting of the SARS-coronavirus Orf7a accessory protein. *Structure* **13:** 75–85 Available at: http://www.ncbi.nlm.nih.gov/pubmed/15642263 [Accessed February 29, 2020]

Netland J, Ferraro D, Pewe L, Olivares H, Gallagher T & Perlman S (2007) Enhancement of Murine Coronavirus Replication by Severe Acute Respiratory Syndrome Coronavirus Protein 6 Requires the N-Terminal Hydrophobic Region but Not C-Terminal Sorting Motifs. *J. Virol.* **81:** 11520–11525

Neuman BW, Kiss G, Kunding AH, Bhella D, Baksh MF, Connelly S, Droese B, Klaus JP, Makino S, Sawicki SG, Siddell SG, Stamou DG, Wilson IA, Kuhn P & Buchmeier MJ (2011) A structural analysis of M protein in coronavirus assembly and morphology. *J. Struct. Biol.* **174:** 11–22 Available at: http://www.ncbi.nlm.nih.gov/pubmed/21130884 [Accessed February 29, 2020]

Oldfield CJ & Dunker AK (2014) Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. *Annu. Rev. Biochem.* **83:** 553–584 Available at: http://www.ncbi.nlm.nih.gov/pubmed/24606139 [Accessed February 29, 2020]

Oostra M, de Haan CAM & Rottier PJM (2007a) The 29-Nucleotide Deletion Present in Human but Not in Animal Severe Acute Respiratory Syndrome Coronaviruses Disrupts the Functional Expression of Open Reading Frame 8. *J. Virol.* **81:** 13876–13888

Oostra M, te Lintelo EG, Deijs M, Verheije MH, Rottier PJM & de Haan CAM (2007b) Localization and Membrane Topology of Coronavirus Nonstructural Protein 4: Involvement of the Early Secretory Pathway in Replication. *J. Virol.* **81:** 12323–12336

Peng K, Radivojac P, Vucetic S, Dunker AK & Obradovic Z (2006) Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics* **7:** 208 Available at: http://www.ncbi.nlm.nih.gov/pubmed/16618368 [Accessed February 29, 2020]

Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK & Obradovic Z (2005) Optimizing long intrinsic disorder predictors with protein evolutionary information. *J. Bioinform. Comput. Biol.* **3:** 35–60

Ponnusamy R, Moll R, Weimar T, Mesters JR & Hilgenfeld R (2008) Variable oligomerization modes in coronavirus non-structural protein 9. *J. Mol. Biol.* **383:** 1081–96 Available at: http://www.ncbi.nlm.nih.gov/pubmed/18694760 [Accessed February 29, 2020]

Rajagopalan K, Mooney SM, Parekh N, Getzenberg RH & Kulkarni P (2011) A majority of the cancer/testis antigens are intrinsically disordered proteins. *J. Cell. Biochem.* **112:** 3256–67 Available at: http://www.ncbi.nlm.nih.gov/pubmed/21748782 [Accessed February 29, 2020]

Ratia K, Saikatendu KS, Santarsiero BD, Barretto N, Baker SC, Stevens RC & Mesecar AD (2006) Severe acute respiratory syndrome coronavirus papain-like protease: structure of a viral deubiquitinating enzyme. *Proc. Natl. Acad. Sci. U. S. A.* **103:** 5717–22 Available at: http://www.ncbi.nlm.nih.gov/pubmed/16581910 [Accessed February 29, 2020]

Ricagno S, Egloff M-P, Ulferts R, Coutard B, Nurizzo D, Campanacci V, Cambillau C, Ziebuhr J & Canard B (2006) Crystal structure and mechanistic determinants of SARS coronavirus nonstructural protein 15 define an endoribonuclease family. *Proc. Natl. Acad. Sci. U. S. A.* **103:** 11892–7 Available at: http://www.ncbi.nlm.nih.gov/pubmed/16882730 [Accessed February 29, 2020]

Robert X & Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42:** W320-4 Available at: http://www.ncbi.nlm.nih.gov/pubmed/24753421 [Accessed February 29, 2020]

Romero P, Obradovic Z, Li X, Garner EC, Brown CJ & Dunker AK (2001) Sequence complexity of disordered protein. *Proteins Struct. Funct. Genet.* **42:** 38–48

Ruch TR & Machamer CE (2012) The coronavirus E protein: assembly and beyond. *Viruses* **4:** 363–82 Available at: http://www.ncbi.nlm.nih.gov/pubmed/22590676 [Accessed February 29, 2020]

Saikatendu KS, Joseph JS, Subramanian V, Neuman BW, Buchmeier MJ, Stevens RC &

Kuhn P (2007) Ribonucleocapsid formation of severe acute respiratory syndrome coronavirus through molecular action of the N-terminal domain of N protein. *J. Virol.* **81:** 3913–21 Available at: http://www.ncbi.nlm.nih.gov/pubmed/17229691 [Accessed February 29, 2020]

Sakai Y, Kawachi K, Terada Y, Omori H, Matsuura Y & Kamitani W (2017) Two-amino acids change in the nsp4 of SARS coronavirus abolishes viral replication. *Virology* **510:** 165–174 Available at: http://www.ncbi.nlm.nih.gov/pubmed/28738245 [Accessed March 7, 2020]

Sawicki SG, Sawicki DL & Siddell SG (2007) A Contemporary View of Coronavirus Transcription. *J. Virol.* **81:** 20–29

Schaecher SR, Mackenzie JM & Pekosz A (2007) The ORF7b Protein of Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) Is Expressed in Virus-Infected Cells and Incorporated into SARS-CoV Particles. *J. Virol.* **81:** 718–731

Sharma K, Åkerström S, Sharma AK, Chow VTK, Teow S, Abrenica B, Booth SA, Booth TF, Mirazimi A & Lal SK (2011) SARS-CoV 9b protein diffuses into nucleus, undergoes active Crm1 mediated nucleocytoplasmic export and triggers apoptosis when retained in the nucleus. *PLoS One* **6:** e19436 Available at: http://www.ncbi.nlm.nih.gov/pubmed/21637748 [Accessed February 29, 2020]

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD & Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7:**

Singh A, Kumar A, Yadav R, Uversky VN & Giri R (2018) Deciphering the dark proteome of Chikungunya virus. *Sci. Rep.* **8:** 5822 Available at: http://www.ncbi.nlm.nih.gov/pubmed/29643398 [Accessed February 29, 2020]

Siu YL, Teoh KT, Lo J, Chan CM, Kien F, Escriou N, Tsao SW, Nicholls JM, Altmeyer R, Peiris JSM, Bruzzone R & Nal B (2008) The M, E, and N Structural Proteins of the Severe Acute Respiratory Syndrome Coronavirus Are Required for Efficient Assembly, Trafficking, and Release of Virus-Like Particles. *J. Virol.* **82:** 11318–11330

Snijder EJ, van der Meer Y, Zevenhoven-Dobbe J, Onderwater JJM, van der Meulen J, Koerten HK & Mommaas AM (2006) Ultrastructure and Origin of Membrane Vesicles Associated with the Severe Acute Respiratory Syndrome Coronavirus Replication Complex. *J. Virol.* **80:** 5927–5940

Song W, Gui M, Wang X & Xiang Y (2018) Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLoS Pathog.* **14:** e1007236 Available at: http://www.ncbi.nlm.nih.gov/pubmed/30102747 [Accessed February 29, 2020]

Sparks JS, Donaldson EF, Lu X, Baric RS & Denison MR (2008) A Novel Mutation in Murine Hepatitis Virus nsp5, the Viral 3C-Like Proteinase, Causes Temperature-Sensitive Defects in Viral Growth and Protein Processing. *J. Virol.* **82:** 5999–6008

Surya W, Samso M & Torres J (2013) Structural and Functional Aspects of Viroporins in Human Respiratory Viruses: Respiratory Syncytial Virus and Coronaviruses. In *Respiratory Disease and Infection - A New Insight* InTech

Tan Y-J (2005) The Severe Acute Respiratory Syndrome (SARS)-coronavirus 3a protein may function as a modulator of the trafficking properties of the spike protein. *Virol. J.* **2:** 5 Available at: http://www.ncbi.nlm.nih.gov/pubmed/15703085 [Accessed February 29, 2020]

Tan Y-J, Teng E, Shen S, Tan THP, Goh P-Y, Fielding BC, Ooi E-E, Tan H-C, Lim SG & Hong W (2004) A Novel Severe Acute Respiratory Syndrome Coronavirus Protein, U274, Is Transported to the Cell Surface and Undergoes Endocytosis. *J. Virol.* **78:**

6723–6734

Tan Y-X, Tan THP, Lee MJ-R, Tham P-Y, Gunalan V, Druce J, Birch C, Catton M, Fu NY, Yu VC & Tan Y-J (2007) Induction of Apoptosis by the Severe Acute Respiratory Syndrome Coronavirus 7a Protein Is Dependent on Its Interaction with the Bcl-XL Protein. *J. Virol.* **81:** 6346–6355

Teoh KT, Siu YL, Chan WL, Schlüter MA, Liu CJ, Peiris JSM, Bruzzone R, Margolis B & Nal B (2010) The SARS coronavirus E protein interacts with PALS1 and alters tight junction formation and epithelial morphogenesis. *Mol. Biol. Cell* **21:** 3838–3852

Thiel V, Ivanov KA, Putics Á, Hertzig T, Schelle B, Bayer S, Weißbrich B, Snijder EJ, Rabenau H, Doerr HW, Gorbalenya AE & Ziebuhr J (2003) Mechanisms and enzymes involved in SARS coronavirus genome expression. *J. Gen. Virol.* **84:** 2305–2315 Available at: http://www.ncbi.nlm.nih.gov/pubmed/12917450 [Accessed March 7, 2020]

Tomar S, Johnston ML, St John SE, Osswald HL, Nyalapatla PR, Paul LN, Ghosh AK, Denison MR & Mesecar AD (2015) Ligand-induced Dimerization of Middle East Respiratory Syndrome (MERS) Coronavirus nsp5 Protease (3CLpro): IMPLICATIONS FOR nsp5 REGULATION AND THE DEVELOPMENT OF ANTIVIRALS. *J. Biol. Chem.* **290:** 19403–22 Available at: http://www.ncbi.nlm.nih.gov/pubmed/26055715 [Accessed March 7, 2020]

Torres J, Wang J, Parthasarathy K & Liu DX (2005) The transmembrane oligomers of coronavirus protein E. *Biophys. J.* **88:** 1283–90 Available at: http://www.ncbi.nlm.nih.gov/pubmed/15713601 [Accessed February 29, 2020]

Tseng Y-T, Chang C-H, Wang S-M, Huang K-J & Wang C-T (2013) Identifying SARS-CoV membrane protein amino acid residues linked to virus-like particle assembly. *PLoS One* **8:** e64013 Available at: http://www.ncbi.nlm.nih.gov/pubmed/23700447 [Accessed February 29, 2020]

Tseng Y-T, Wang S-M, Huang K-J, Lee AI-R, Chiang C-C & Wang C-T (2010) Self-assembly of severe acute respiratory syndrome coronavirus membrane protein. *J. Biol. Chem.* **285:** 12862–72 Available at: http://www.ncbi.nlm.nih.gov/pubmed/20154085 [Accessed February 29, 2020]

Ujike M & Taguchi F (2015) Incorporation of spike and membrane glycoproteins into coronavirus virions. *Viruses* **7:** 1700–1725

Uversky VN, Gillespie JR & Fink AL (2000) Why are 'natively unfolded' proteins unstructured under physiologic conditions? *Proteins Struct. Funct. Genet.* **41:** 415–427

Uversky VN, Oldfield CJ & Dunker AK (2005) Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J. Mol. Recognit.* **18:** 343–84 Available at: http://www.ncbi.nlm.nih.gov/pubmed/16094605 [Accessed February 29, 2020]

Varshney B & Lal SK (2011) SARS-CoV accessory protein 3b induces AP-1 transcriptional activity through activation of JNK and ERK pathways. *Biochemistry* **50:** 5419–25 Available at: http://www.ncbi.nlm.nih.gov/pubmed/21561061 [Accessed February 29, 2020]

te Velthuis AJW, van den Worm SHE & Snijder EJ (2012) The SARS-coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both de novo initiation and primer extension. *Nucleic Acids Res.* **40:** 1737–47 Available at: http://www.ncbi.nlm.nih.gov/pubmed/22039154 [Accessed February 29, 2020]

Voss D, Pfefferle S, Drosten C, Stevermann L, Traggiai E, Lanzavecchia A & Becker S (2009) Studies on membrane topology, N-glycosylation and functionality of SARS-CoV membrane protein. *Virol. J.* **6:** 79 Available at: http://www.ncbi.nlm.nih.gov/pubmed/19534833 [Accessed February 29, 2020]

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. *J. Mol. Biol.*

337: 635–645

Woo PCY, Lau SKP, Lam CSF, Lau CCY, Tsang AKL, Lau JHN, Bai R, Teng JLL, Tsang CCC, Wang M, Zheng B-J, Chan K-H & Yuen K-Y (2012) Discovery of Seven Novel Mammalian and Avian Coronaviruses in the Genus Deltacoronavirus Supports Bat Coronaviruses as the Gene Source of Alphacoronavirus and Betacoronavirus and Avian Coronaviruses as the Gene Source of Gammacoronavirus and Deltacoronavi. *J. Virol.* **86:** 3995–4008

Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O, Graham BS & McLellan JS (2020) Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* Available at: http://www.ncbi.nlm.nih.gov/pubmed/32075877 [Accessed February 29, 2020]

Wright PE & Dyson HJ (1999) Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **293:** 321–331

Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P, Meng J, Zhu Z, Zhang Z, Wang J, Sheng J, Quan L, Xia Z, Tan W, Cheng G & Jiang T (2020a) Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. *Cell Host Microbe* Available at: http://www.ncbi.nlm.nih.gov/pubmed/32035028 [Accessed March 7, 2020]

Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, Hu Y, Tao Z-W, Tian J-H, Pei Y-Y, Yuan M-L, Zhang Y-LY-Z, Dai F-H, Liu Y, Wang Q-M, Zheng J-J, Xu L, Holmes EC & Zhang Y-LY-Z (2020b) A new coronavirus associated with human respiratory disease in China. *Nature*

Xu K, Zheng B-J, Zeng R, Lu W, Lin Y-P, Xue L, Li L, Yang L-L, Xu C, Dai J, Wang F, Li Q, Dong Q-X, Yang R-F, Wu J-R & Sun B (2009) Severe acute respiratory syndrome coronavirus accessory protein 9b is a virion-associated protein. *Virology* **388:** 279–85 Available at: http://www.ncbi.nlm.nih.gov/pubmed/19394665 [Accessed February 29, 2020]

Xue B, Dunbrack RL, Williams RW, Dunker AK & Uversky VN (2010a) PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim. Biophys. Acta* **1804:** 996–1010 Available at: http://www.ncbi.nlm.nih.gov/pubmed/20100603 [Accessed February 29, 2020]

Xue B, W. Williams R, J. Oldfield C, Kian-Meng Goh G, Keith Dunker A & N. Uversky V (2010b) Viral Disorder or Disordered Viruses: Do Viral Proteins Possess Unique Features? *Protein Pept. Lett.* **17:** 932–951

Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, Wu Y, Zhang L, Yu Z, Fang M, Yu T, Wang Y, Pan S, Zou X, Yuan S & Shang Y (2020) Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *Lancet. Respir. Med.* Available at: http://www.ncbi.nlm.nih.gov/pubmed/32105632 [Accessed February 29, 2020]

Yu C-J, Chen Y-C, Hsiao C-H, Kuo T-C, Chang SC, Lu C-Y, Wei W-C, Lee C-H, Huang L-M, Chang M-F, Ho H-N & Lee F-JS (2004) Identification of a novel protein 3a from severe acute respiratory syndrome coronavirus. *FEBS Lett.* **565:** 111–6 Available at: http://www.ncbi.nlm.nih.gov/pubmed/15135062 [Accessed February 29, 2020]

Yu I-M, Oldham ML, Zhang J & Chen J (2006) Crystal structure of the severe acute respiratory syndrome (SARS) coronavirus nucleocapsid protein dimerization domain reveals evolutionary linkage between corona- and arteriviridae. *J. Biol. Chem.* **281:** 17134–9 Available at: http://www.ncbi.nlm.nih.gov/pubmed/16627473 [Accessed February 29, 2020]

Yuan X, Li J, Shan Y, Yang Z, Zhao Z, Chen B, Yao Z, Dong B, Wang S, Chen J & Cong Y (2005a) Subcellular localization and membrane association of SARS-CoV 3a protein.

Virus Res. **109:** 191–202 Available at: http://www.ncbi.nlm.nih.gov/pubmed/15763150 [Accessed February 29, 2020]

Yuan X, Shan Y, Yao Z, Li J, Zhao Z, Chen J & Cong Y (2006) Mitochondrial location of severe acute respiratory syndrome coronavirus 3b protein. *Mol. Cells* **21:** 186–91 Available at: http://www.ncbi.nlm.nih.gov/pubmed/16682811 [Accessed February 29, 2020]

Yuan X, Shan Y, Zhao Z, Chen J & Cong Y (2005b) G0/G1 arrest and apoptosis induced by SARS-CoV 3b protein in transfected cells. *Virol. J.* **2:** 66 Available at: http://www.ncbi.nlm.nih.gov/pubmed/16107218 [Accessed February 29, 2020]

Yuan X, Yao Z, Shan Y, Chen B, Yang Z, Wu J, Zhao Z, Chen J & Cong Y (2005c) Nucleolar localization of non-structural protein 3b, a protein specifically encoded by the severe acute respiratory syndrome coronavirus. *Virus Res.* **114:** 70–9 Available at: http://www.ncbi.nlm.nih.gov/pubmed/16046244 [Accessed February 29, 2020]

Zhai Y, Sun F, Li X, Pang H, Xu X, Bartlam M & Rao Z (2005) Insights into SARS-CoV transcription and replication from the structure of the nsp7-nsp8 hexadecamer. *Nat. Struct. Mol. Biol.* **12:** 980–6 Available at: http://www.ncbi.nlm.nih.gov/pubmed/16228002 [Accessed March 7, 2020]

Zhou H, Ferraro D, Zhao J, Hussain S, Shao J, Trujillo J, Netland J, Gallagher T & Perlman S (2010) The N-Terminal Region of Severe Acute Respiratory Syndrome Coronavirus Protein 6 Induces Membrane Rearrangement and Enhances Virus Replication. *J. Virol.* **84:** 3542–3551