

The article by Dhulipala and Uversky examines the predicted disorder in the proteome of RABV and its human proteome interaction network. The conclusions claimed in this article are not supported, as a logical link between a predicted disorder profile and function is never made. Additionally, in many places the citations used to suggest a particular function are inappropriate as the cited article does not refer to said function. Please see below for additional details.

Major comments:

The abstract states, "The study suggests that the high levels of intrinsic disorder in the phosphoprotein (P-protein) and nucleoprotein (N-protein) allow them to participate in creation of the Negri bodies and help this virus suppress the antiviral immune response in the host cells. Additionally, the study suggests that there is a link between disorder in the matrix (M) protein and the modulation of viral transcription. The disordered regions in the M protein have a possible role in initiating viral budding within the cell." However, just showing that a protein has x% disorder does not imply that the disorder regions are important for each of the different functions laid out. Thus, this is misleading because the study does not provide a direct link to disorder playing a specific role in any of these functions. For a link between disorder and function to be made more than just a predicted disorder profile has to be shown. Thus, such statements throughout the article need to be greatly toned down in terms of this study providing evidence for functional relevance.

On pg. 7 it states, "Additionally, this high disorder content suggests that the mechanism of action for the P-protein to suppress the Type-I interferon-mediated immune response may be to use its disordered status for interaction with the STAT proteins of the cells... The IDPRs that were found in this protein corresponded to ligand binding sites, which supports the idea that the IDPRs interact with the STAT1 and STAT2 proteins". How does having high disorder imply that disorder is important for a specific interaction with another protein, particularly the STAT proteins? How does the IDP having a general "ligand binding site" imply that the IDP interacts with STAT1 and STAT2? These statements need to be explained and unless there is a specific feature the authors can point out this seems like too big of a jump. Furthermore, the citation (Wiltzer et al.) used to support an interaction between the P-protein and STAT proteins shows that the interaction is driven by W-hole residues in the folded region of the P-protein and $\Delta C30$, which removes this folded domain, eliminates interactions between the P-protein and the STAT proteins. Thus, evidence supporting that the IDRs are important for interactions with STAT proteins is severely lacking.

On pg. 9 FuzDrop scores are given for P-protein isoforms, yet the reader is given no context about what these scores mean. Does 0.52 mean it has a 50% chance of undergoing LLPS? For these scores to have meaning for the reader some scale information is needed such as what is the average score for proteins that are known to undergo LLPS and where do the P-protein scores fall on the distribution of scores of a large set of proteins? Do these values actually support the claim that the protein "has a strong LLPS potential"?

On pg. 9 it states, “the MATRX_RABVP protein and shows that a section in the middle of the protein, spanning roughly from residue 50 to residue 175, displays low disorder content and is likely to represent the structural domain of the protein. Curiously, region 115-151, which is essential for the glycoprotein binding includes an IDPR (residues 129-141), indicating that intrinsic disorder of this region can contribute to its interactability.” These two sentences appear contradictory. The claim is the region defined by residues 50-175 is ordered but then it’s the “disordered” part (129-141) that is important for binding even though the disorder score for this region is ~0.5 and does not make up a majority of the region 115-151.

On pg. 12 it states, “The N-protein also functions to prevent activation of the RIG-I mediated antiviral response [19,20]. The IDPR found in the N-protein may serve to aid in suppressing the antiviral response.” Again there is a lack of logic flow to connect predicted disorder regions to a listed function.

In the abstract it is stated, “The study suggests that the high levels of intrinsic disorder in the phosphoprotein (P-protein) and nucleoprotein (N-protein) allow them to participate in creation of the Negri bodies and help this virus suppress the antiviral immune response in the host cells”. Yet on pg. 12 it states that the N-protein has a low probability of undergoing LLPS according to FuzDrop. Thus, again there is a lack of connection between the presence of predicted disorder, LLPS, and Negri body formation.

On pg. 14 there is a lack of logic flow for the statement, “The IDPR in the G-protein likely serves to facilitate endocytosis into the cell through interacting with cell recognition molecules.” Also it is unclear why Zhang et al., is cited for this statement given endocytosis is not mentioned in Zhang et al.

On pg. 14 it states, “The IDPRs likely define the lower fidelity of the polymerase, which results in a high mutation rate and therefore higher flexibility in the ability of the virus to adapt to host defenses [7].” Again there is no logical connection between predicted disorder and function, and the citation (Kouznetsov et al.) appears to just be about the N protein.

On pg. 17 it states, “This revealed a very high level of disorder in the majority of the proteins from this dataset, with the entire set being characterized by the mean PPIDR of 41.6 ± 20.9 % (as evaluated using the outputs of the PONDR® VSL2 predictor, which is one of the most accurate stand-alone disorder predictors). Supplementary Figures S1, S2, S3, S4, and S5 654 show that all these proteins contain multiple IDPRs of various length. Many proteins contain multiple MoRFs, and almost all human proteins in this dataset are densely decorated by a multitude of different PTMs. These observations indicate that intrinsic disorder in these proteins is related to their functionality, playing a role in their binding promiscuity, as evidenced by dense PPI networks centered at these proteins.” There is no reference to whether these observations are statistically significant. This dataset could be directly compared to the full human proteome to determine if the mean PPIDR is significant. And even if the mean PPIDR is greater than the human

proteome it does not automatically follow that “intrinsic disorder in [this] these proteins is related to their functionality”.

On pg. 17 it states, “In fact, Figure 7A shows there is no a single protein in this dataset that could be classified as mostly ordered, whereas 25 proteins (67.6%) are expected to be mostly disordered (i.e., their PPIDR exceeds 30%).” Why is 30% defined as “mostly disordered”. PONDR defines disorder as anything greater than or equal to 0.5. This is also the line cutoff used in all of the previous disorder predictor plots in this paper. Thus, according to this definition many of the proteins are “mostly” ordered. The reasoning for 30% needs to be addressed and why it is different than what PONDR suggests. For instance, PONDR VSL2 (%) which is defined in the figure caption as the percent of “residues with disorder scores above 0.5” is less than 50% for many proteins which implies that those proteins are mostly ordered. Furthermore, the same symbols and colors are repeated in Figure 7 making it impossible to distinguish between certain proteins.

On pg. 17 it states, “Figure 7B shows that quadrant Q1 (bottom left corner) contains 21 proteins that are predicted to be ordered by both predictors; quadrant Q2 (bottom right corner) includes 10 proteins, which are predicted to be ordered/compact by CH-plot and disordered by CDF (i.e., it contains either molten globular proteins, which are compact, but without unique 3D structures, or hybrid proteins containing comparable levels of ordered and disordered residues); quadrant Q3 (top right corner) includes 4 highly disordered (native coils or native pre-molten globules), which are predicted as disordered by both predictors. Finally, one proteins in quadrant Q4 (top left corner) is predicted to be disordered by CH-plot and ordered by CDF-plot. Therefore, 16 human proteins interacting with the RABV proteins are predicted as containing very noticeable levels of disorder” This description does not match the figure. Particularly, the bottom right has the most proteins, but is said to only have 10 proteins. Also where does the number 16 come from when the quadrants contain 21, 10, 4, and 1? Also, if the 21 proteins are predicted to be ordered by both predictors, then this whole plot is in contrast to what was stated about these proteins in Figure 7A, which claimed “no single protein in this dataset could be classified as mostly ordered”.

Minor comments:

On pg. 9 there is no citation for the statement, “Curiously, region 115-151, which is essential for the glycoprotein binding includes an IDPR (residues 129-141), indicating that intrinsic disorder of this region can contribute to its interactability.”

On pg. 10 one of the 3Bs should be 3A in the sentence, “Comparison of the Figures 3B and 3B illustrates a remarkable similarity of the per-residue intrinsic disorder predispositions of the Lagos...”.

On pg. 11 it states, “The N-protein has been shown to form a complex with the P-protein during replication and encapsulate the viral genomic RNA to protect it against nucleases. It binds through the N-terminus of the protein, which is highly intrinsically disordered.” However, there are no citations to support these claims. Also only the first ~10 residues

of the N-terminus are predicted to be disordered so is this the region that binds to the P-protein or a region larger than this? And then in the next sentence it implies the C-terminus is the actual region that interactions with the P-protein. Also the citation that goes with this sentence (Luo et al.) does not show what the sentence claims. This text needs to be cited correctly and clarified.

On pg. 12 it states, "The wide curvature of the C terminus of the protein, which is highly intrinsically disordered, makes it able to more tightly cover bound RNA." What does "wide curvature" mean in terms of region that adopts heterogeneous conformations? The references to curvature from Luo et al. seems to be in terms of the structured domains. The meaning of this statement needs to be clarified.

On pg. 12 it states, "The C-terminal domain of the G protein is essential for trimer stability". Please clarify what the "C-terminal domain" refers to. Is it residues 258-505?

On pg. 12 it states, "The most promising IDPR is the C-terminal IDPR (residues 486-524), which corresponds to the intravirion domain of this protein that is engaged in interaction with the matrix protein." This statement has no citation.

Line 527 has two Figure 5Bs.

There are several typos on pg. 16 including lines 582 and 605.

It is not stated how "mostly disordered" is calculated for Figure 9 and there is no color bar. If the authors are using 30%, then this value needs to be explained because 30% does not generally imply the value is high.