Article

Looking at the Pathogenesis of the Rabies Lyssavirus Strain Pasteur Vaccins through a Prism of the Disorder-Based Bioinformatics

Surya Dhulipala [1] and Vladimir N. Uversky [1,2,3,*]

1 Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA
2 USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA
3 Protein Research Group, Institute for Biological Instrumentation of the Russian Academy of Sciences, Federal Research Center "Pushchino Scientific Center for Biological Research of the Russian Academy of Sciences", 142290 Pushchino, Moscow Region, Russia
* Correspondence: vuversky@usf.edu

Abstract: Rabies is a neurological disease that causes between 40,000 and 70,000 deaths every year. Once a rabies patient has become symptomatic, there is no effective treatment for the illness, and in unvaccinated individuals, the case-fatality rate of rabies is close to 100%. French scientists Louis Pasteur and Émile Roux developed the first vaccine for rabies in 1885. If administered before the virus reaches the brain, the modern rabies vaccine imparts long-lasting immunity to the virus and saves more than 250,000 people every year. However, the rabies virus can suppress the host's immune response once it has entered the cells of the brain, making death likely. This study aimed to make use of disorder-based proteomics and bioinformatics to determine the potential impact that intrinsically disordered protein regions (IDPRs) in the proteome of the rabies virus might have on the infectivity and lethality of the disease. This study used the proteome of the Rabies lyssavirus (RABV) strain Pasteur Vaccins (PV), one of the best-understood strains due to its use in the first rabies vaccine, as a model. The data reported in this study are in line with the hypothesis that high levels of intrinsic disorder in the phosphoprotein (P-protein) and nucleoprotein (N-protein) allow them to participate in the creation of Negri bodies and might help this virus to suppress the antiviral immune response in the host cells. Additionally, the study suggests that there could be a link between disorder in the matrix (M) protein and the modulation of viral transcription. The disordered regions in the M-protein might have a possible role in initiating viral budding within the cell. Furthermore, we checked the prevalence of functional disorder in a set of 37 host proteins directly involved in the interaction with the RABV proteins. The hope is that these new insights will aid in the development of treatments for rabies that are effective after infection.

Keywords: rabies; intrinsic disorder; intrinsically disordered protein; intrinsically disordered protein region; protein–protein interaction

1. Introduction

Rabies lyssavirus is a bullet-shaped, negative-sense, unsegmented, single-stranded RNA virus of the *Rhabdoviridae* family. There are 10 viruses in the Rabies serogroup, but most are not pathogenic to humans. Rabies lyssavirus and Australian bat lyssavirus are the only two rhabdoviruses that have been known to cause disease in humans [1]. The rabies virus (RABV) is a zoonotic neurotropic virus that causes fatal neurological symptoms in almost all mammals and is spread through the bite of an infected mammal. Rabies disease causes between 40,000 and 70,000 deaths every year worldwide. Once a rabies patient has become symptomatic, there is no effective treatment for the illness. In fact, in unvaccinated

individuals, the case-fatality rate of rabies is close to 100% [2]. French scientists Louis Pasteur and Émile Roux developed the first vaccine for rabies in 1885. If administered before the virus reaches the brain, the modern rabies vaccine imparts long-lasting immunity to the virus. It saves more than 250,000 people every year. RABV is able to suppress the host's immune response once it has entered the cells of the brain, making death likely if the vaccine is not administered [3]. The virus works by infecting muscle cells through saliva in bites before traveling to the brain. The virus travels through neuromuscular junctions and neural pathways using axoplasmic transport. Once the virus reaches the brain, it triggers endocytosis into neurons. It then undergoes transcription and replication by polymerase shuttering [4]. The virus suppresses the antiviral response and infects more cells before eventually being released and transported to the salivary glands. The high neuroinvasiveness of RABV has been attributed to the ability of the virus to evade immune responses through various means and to conserve the structures of neurons [3].

All RABVs encode for five major proteins, as well as four more isoforms of the phosphoprotein that are generated through alternative initiation. The major proteins are the phosphoprotein (P-protein), matrix protein (M-protein), glycoprotein (G-protein), nucleoprotein (N-protein), and polymerase, or the large protein (L-protein). A total of five phosphoproteins can be made through alternative initiation [5]. The RNA genome of the virus is roughly 11 kb in size. A graphical representation of the virus can be seen in Figure 1. The P-protein is bound tightly around the RNA of the virus to create the RNP (ribonucleoprotein) core, and the RNP condenses with the N- and L-proteins to create the helical nucleocapsid of the virus. The N-protein homomultimerizes to form the nucleocapsid and binds with the P-protein [6]. The P-protein forms a homotrimer when phosphorylated and bound to the L-protein, which positions the L-protein over the template RNA strands [7]. The M-protein forms a homomultimer between the nucleocapsid core and the glycoprotein spikes that surround the virus to connect the two. The G-protein spikes, which are homotrimers of the G-protein, form the sole ligand for the cellular receptor. Anti-rabies antibodies target the G-protein homotrimers [8,9].
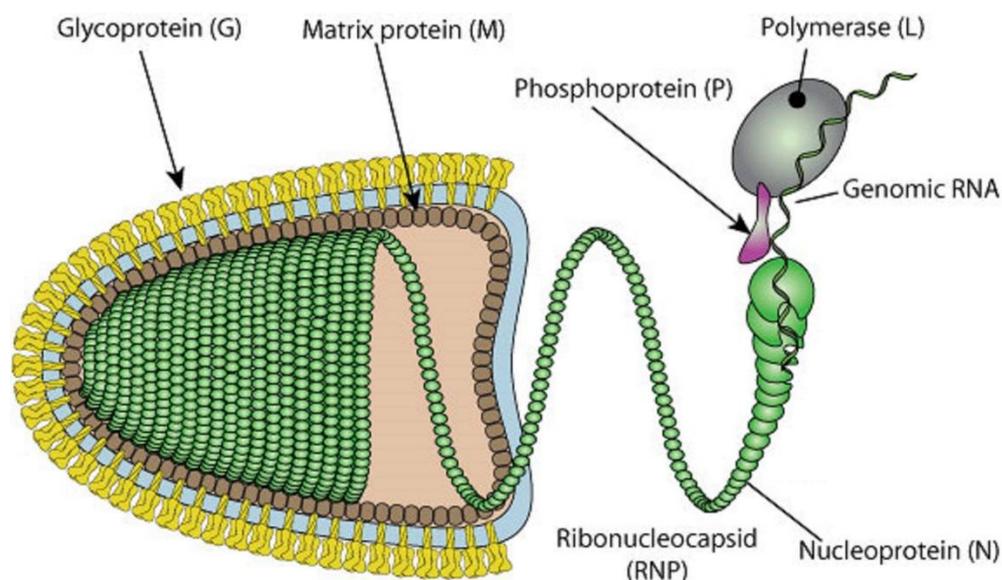

Figure 1. Structural features of RABV. In mature RABV, the nucleoprotein, phosphoprotein, and viral polymerase envelop the genomic RNA in a structure known as the ribonucleocapsid (RNP). The matrix protein surrounds the RNP and determines the shape of the virus. The matrix protein also anchors the glycoprotein to the envelope [10] (original source of the image: Philippe Le Mercier, SIB Swiss Institute of Bioinformatics).

Once the virus reaches the cell, the G-protein initiates endocytosis. The low pH inside the endosome initiates the fusion of the membrane with the endosome, resulting in the

release of the nucleocapsid into the cell. The P-protein acts as a cofactor to the L-protein, creating the functional viral polymerase [11]. The M-protein acts as a bridge between the plasma membrane and the generated nucleocapsids, providing a crucial role in virus budding [3,12,13]. The condensing of the nucleocapsid and the final shape is mediated by the M-protein [3].

The P-protein has an important role in the infection and propagation of RABV. It binds to the dynein light-chain proteins DYNLL1 and DYNLL2 in humans, allowing the virus to be transported along the axon to the nucleus. During transcription, alternative initiation using a ribosomal leaky scanning mechanism can create five variations of the P-protein, termed P1–5 [14]. This enables higher flexibility in the action of the P-protein. It interacts with the ribosomal protein L9 during the initial stages of RABV infection to inhibit RABV transcription during this period, controlling viral replication [15]. The P-protein also acts as an antagonist to STAT1 and STAT2 proteins within the cell. STAT proteins, which are activated by type-I-interferon receptors in the type-I-interferon-mediated innate immune response, are cytoplasmic signal transducers and transcription activators associated with the immune response [16,17]. The P-protein exploits microtubule processes within the cell to transport the virus and force STAT1 to use microtubule-inhibited mechanisms, suppressing the nuclear transport of STAT and signaling [18]. STAT proteins are important for the establishment of the antiviral state in the cell. The STAT-targeting feature of interferon antagonists is considered a determining factor in the pathogenicity of the virus. Inhibition of the interaction between the P-protein and STAT has been shown to severely reduce lethality and increase the immune response to RABV. The N-protein forms a complex with the P-protein to encapsulate the RNA genome of the virus and protect it from nuclease activity and phosphorylation [6]. The N-protein has been shown to help suppress the type-3-interferon-mediated immune system by evading the RIG-I-mediated antiviral response [19,20]. The L-protein, due to its activity as a viral polymerase, plays multiple enzymatic roles in the synthesis and processing of viral RNA [21]. It acts as a low-fidelity polymerase, resulting in a high rate of mutations in RABV.

The M-protein has been shown to be significant in regulating viral budding. It has been shown to bind to the plasma membrane even in the absence of other viral proteins, which suggests that the protein acts to connect the plasma membrane to the nucleocapsid. The M-protein has been shown to be able to induce vesicle budding without necessitating interactions with other proteins [3,12]. The infection of a cell with an M-protein-deficient RABV mutant was observed to result in smaller numbers of rod-shaped or round viral particles rather than the bullet-shaped viruses that are standard with mature RABV [13]. The M-protein has been shown to inhibit viral transcription in many viruses, including RABV, suggesting that it has a highly regulatory role in transcription and replication [12,22]. The M-protein has been shown to interact with RelAp43, a protein in the NF-*κ*B family of proteins, which results in the further suppression of the antiviral immune response [23]. The G-protein is associated with the acceleration of the budding efficiency of the virus [12,13]. The G-protein also drastically increases pathogenicity because, as the only surface protein in the virus, it induces innate immune responses by binding to immune receptors [24]. This binding is shown to promote effective virus uptake, which drastically increases virulence.

As is typical for other viruses, all RABV proteins are multifunctional. Such multifunctionality is crucial, as it allows a very small set of viral proteins to manage and control each and any aspect of the viral "life", from entry to replication to the formation and exit of new infectious particles, and regulate each and any aspect of virus interaction with the host. In fact, a typical protein repertoire of a typical virus includes a minimal set of specific structural proteins that are crucial for the viral particle assembly and a set of non-structural proteins needed for the hijacking of many functional pathways of the host cell. This is why many viral proteins are multifunctional. They are typically engaged in numerous interactions with various host cell components. In the protein universe, multifunctionality and binding promiscuity are typically associated with protein intrinsic disorder. This directly follows from the recognition that many protein functions do not need a unique 3D structure [25–33]

and that such structure-less intrinsically disordered proteins (IDPs) and intrinsically disordered protein regions (IDPRs) are commonly found in various proteomes [34–38], where they are involved in regulation, signaling, and control pathways [27,30,32,39–41], thereby possessing functions that complement the functional repertoire of traditional ordered proteins [42–47]. IDPs/IDPRs are often involved in various human diseases [48,49]. They have complex mosaic structures and show remarkable multi-level spatiotemporal heterogeneity, existing as dynamic conformational ensembles [25,27,31,40,42,50,51] where different parts of a protein can be (dis)ordered to different degrees [40,52,53]. Importantly, these differently (dis)ordered pieces of the protein structural mosaic might have well-defined and specific functions [53]. Therefore, IDPs/IDPRs are structurally and functionally heterogeneous complex systems whose functionality is described in terms of the protein structure–function continuum [53,54], where the structural and functional diversification of a protein is defined by several factors determining the capability of a single gene to encode a set of distinct protein molecules, known as proteoforms [55]. This is achieved at several levels by altering the chemical structure of the proteinaceous product(s) of a given gene via allelic variations at the DNA level (utilizing several specific means, such as single or multiple point mutations, indels, and SNPs) by alternative splicing and other pre-translational mechanisms affecting mRNA, via a broad arsenal of countless post-translational modifications (PTMs) of a polypeptide chain, by the presence of intrinsic disorder, or by structural alterations induced by functioning [54].

Importantly, based on computational analyses of the abundance of intrinsic disorder in various organisms, it has been concluded that the proteomes of viruses have the largest variability in the content of disordered residues in comparison with all other kingdoms of life [37,38]. The abundance and functional importance of IDPs/IDPRs have been systematically investigated for human papillomaviruses (HPVs) [56,57], human immunodeficiency virus 1 (HIV-1) [58], human hepatitis C virus (HCV) [59,60], Dengue virus [61], rotavirus [62], human respiratory syncytial virus [63], Zika virus [64,65], Chikungunya virus [66], Alkhurma virus (ALKV) [67], Japanese encephalitis virus [68], and SARS-CoV-2, human SARS, and bat SARS-like coronaviruses [69]. These studies suggested that the presence of IDPRs in viral proteins is crucial for their functionality and represents an important means of the overall enhancement of viral propagation during the virus life cycle. To the best of our knowledge, there is no similar analysis of the disorder status of the RABV proteome. The aim of this study was to fill this gap by conducting a comprehensive computational analysis of the penetrance of intrinsic disorder in the proteins of the Rabies lyssavirus strain Pasteur Vaccins proteome.

2. Materials and Methods

The Universal Protein Resource (UniProt) is an annotated database of protein sequences [70]. Entries may be manually annotated with information extracted from the literature and evaluated by computational analysis (Swiss-Prot) or computationally analyzed (TrEMBL). The Rabies lyssavirus strain Pasteur Vaccins proteome is a Swiss-Prot annotated entry. Each entry contains information, including but not limited to the taxonomy, molecular function and included biological processes, subcellular location, potential modifications, pathology, interactions, and structures. Of these, the amino acid sequences and the structures are the most valuable resources for the purposes of studying intrinsic disorder in the virus. A set of 37 human proteins involved in interactions with the RABV proteins was assembled through a literature search, with the majority of data being retrieved from [71]. Amino acid sequences and basic disorder-related features of these proteins are provided in Supplementary Figures S1–S5.

The amino acid sequences can be analyzed through various methods to identify regions in the protein that have a predisposition towards intrinsic disorder. Through a comparison of these regions of intrinsic disorder and the function of the protein, an analysis can be made of the disorder-based functionality of the protein. Determining the structure of the protein is important for mapping the intrinsic regions of disorder to regions on the

protein. Intrinsically disordered regions are flexible. As a result, these regions are not recorded in the crystal structure of the protein. Regions of the protein that do not show up in the crystal structure are indicative of intrinsic disorder [72].

For this study, amino acid FASTA sequences for the Rabies lyssavirus PV proteome were gathered through UniProt. The sequences were run through a series of disorder prediction tools to generate an estimate of the intrinsic disorder of each residue. These predictions of intrinsic disorder were averaged to form an overall prediction of intrinsic disorder for that residue. Table 1 shows a summary of these predictions, including the UniProt entry ID of each protein analyzed. The table also displays the protein length and the length of the longest region of disorder within the protein. Overall disorder was calculated based on the incidence of regions of high disorder. The FASTA sequences used in this study are reproduced in Supplementary Materials. The underlined sections of each amino acid sequence correspond to regions of high disorder within the protein. The amino acid sequences presented here were taken from UniProt [70]. Regions of high intrinsic disorder (intrinsic disorder of more than 30%) are shown using bold underlined text. The per-residue intrinsic disorder propensity for each protein is calculated by taking the averages of the PONDR® VL3, PONDR® VLS2, PONDR® VLXT, PONDR® FIT, IUPred_Short, and IUPred_Long predictors.

Table 1. Proteins in the rabies PV proteome and their evaluated percentages of intrinsically disordered residues as determined by the averaged predicted disorder.

Protein	UniProt Entry ID [70]	Protein Length (Residues)	Longest Disordered Region (Residues)	Percent of Disordered Residues
P (Phosphoprotein)	P06747	297	87	67.3%
M (Matrix protein)	P08671	202	54	43%
N (Nucleoprotein)	P06025	450	93	30.6%
G (Glycoprotein)	P08667	524	49	27%
L (Large protein)	P11213	2142	105	23%

Predictor of Natural Disordered Regions (PONDR®) is meta-prediction software that can analyze an amino acid FASTA sequence on a per-residue basis to predict regions of intrinsic disorder. This study made use of the PONDR®VLXT, PONDR®VL3, PONDR®VLS2, and PONDR® FIT meta-predictors from the PONDR family, as well as IUPred_Short and IUPred_Long [73–75].

PONDR®VLXT works by applying three different neural networks: one for each terminal end of an intrinsically disordered sequence and one for the internal region of the sequence. Each network uses a specific dataset that contains only the amino acid residues that are present in that region. The result of the predictor is an average of the results of the three networks. Transitions between the prediction networks work by averaging the predictors in a short region of overlap at the boundary between the two. PONDR® VLXT is useful for predicting short regions of disorder but underestimates the occurrence of long disordered regions. PONDR® VL3 works by running the residue through ten neural networks and selecting the final prediction by taking the simple majority vote of the predictions. This meta-predictor is known to be useful for predicting longer regions of intrinsic disorder. PONDR® VSL2 combines neural network predictors for short and long disordered regions. The networks are trained using sequences of specific lengths, and the final prediction is a weighted average of the predictions for each length. Because it combines both short and long disordered regions, it is considered the most accurate predictor of the three [73–75].

IUPred works from the assumption that globular and structured proteins have higher numbers of effective inter-residue interactions than disordered proteins do, which means that they have negative free-energy. Structured proteins have lower free-energy estimates compared to disordered proteins. The IUPred meta-predictor is able to use this biophysics-based approach to estimate disorder by calculating the pairwise free-energy of the se-

quence [74,75]. IUPred Long predicts global structural disorder, or disorder in regions with more than 30 consecutive residues. IUPred Short is useful for predicting short disordered regions, such as the region corresponding to the missing residues in the X-ray structure of a largely globular protein. ANCHOR2 is used to predict context-dependent intrinsic disorder. Context-dependent intrinsic disorder may occur when the binding region of an IDPR is able to interact specifically with a globular protein. When bound, these regions adopt an ordered structure. Context-dependent intrinsic disorder may also occur when the change in disorder is due to a change in the redox state. These regions may change their disorder depending on their localization relative to the cell. For all query proteins, the presence of such context-dependent disordered regions, disorder-based binding regions, and molecular recognition features (MoRFs), i.e., disordered regions that fold upon their interaction with partners, was analyzed by the ANCHOR algorithm [76,77].

A recently designed computational platform, RIDAO (Rapid Intrinsic Disorder Analysis Online), was used to obtain the intrinsic-disorder-related characteristics of the query proteins [78]. This tool aggregates the results from a number of well-known disorder predictors: PONDR® VLXT [79], PONDR® VL3 [80], PONDR® VLS2B [81], PONDR® FIT [73], IUPred2 (Short), and IUPred2 (Long) [74,82]. Furthermore, RIDAO provides the mean disorder profile (MDP), along with the standard errors. It also performs CH-CDF (charge-hydropathy–cumulative distribution function) analysis of the query proteins [83–85] and yields data for ΔCH-ΔCDF plots, which enables rapid discrimination between flavors of disorder [86].

The outputs of the per-residue predictors were averaged, and proteins were grouped based on their percentages of predicted intrinsically disordered residues (PPIDR) using accepted classification criteria [87]. Proteins with an average content of intrinsically disordered residues below 10% are considered ordered or mostly ordered. Proteins containing between 10% and 30% predicted disordered residues are considered moderately disordered. Proteins containing more than 30% predicted disordered residues are considered highly disordered [87].

Complementary disorder evaluations, together with important disorder-related functional information, were retrieved from the D^2P^2 database (http://d2p2.pro/, accessed on 15 August 2022) [88], which is a database of predicted disorder for a large library of proteins from completely sequenced genomes [88]. The D^2P^2 database uses the outputs of IUPred [74,82], PONDR® VLXT [79], PrDOS [89], PONDR® VSL2B [81], PV2 [88], and ESpritz [90]. The visual console of D^2P^2 displays 9 colored bars representing the location of disordered regions, as predicted by these different disorder predictors. In the middle of the D^2P^2 plots, the blue–green–white bar shows the predicted disorder agreement between the nine disorder predictors (IUPred, PONDR® VLXT, PONDR® VSL2, PrDOS, PV2, and ESpritz), with blue and green parts corresponding to disordered regions by consensus. Above the disorder consensus bar are two lines with colored and numbered bars that show the positions of the predicted (mostly structured) SCOP domains [91,92] using the SUPERFAMILY predictor [93]. The yellow zigzagged bar shows the location of the predicted disorder-based binding sites (MoRF regions) identified by the ANCHOR algorithm [76], whereas differently colored circles at the bottom of the plot show the locations of various PTMs assigned using the outputs of the PhosphoSitePlus platform [94], which is a comprehensive resource for experimentally determined post-translational modifications.

Information on the interactability of human proteins that interact with RABV proteins was retrieved using Search Tool for the Retrieval of Interacting Genes (STRING, http://string-db.org/, accessed on 15 August 2022). STRING generates a network of protein–protein interactions based on predicted and experimentally validated information on the interaction partners of a protein of interest [95]. In the corresponding network, the nodes correspond to proteins, whereas the edges show predicted or known functional associations. Seven types of evidence are used to build the corresponding network, where they are indicated by differently colored lines: a green line represents neighborhood evidence; a red line represents the presence of fusion evidence; a purple line represents experimental

evidence; a blue line represents co-occurrence evidence; a light blue line represents database evidence; a yellow line represents text mining evidence; and a black line represents co-expression evidence [95]. In this study, STRING was utilized in three different modes: to create PPI networks centered on individual human proteins, to generate the internal network of protein–protein interactions (PPIs) among the human proteins involved in interactions with the RABV proteins, and to build a PPI network centered on the entire set.

The propensity of the RABV proteins to undergo liquid–liquid phase separation was evaluated by FuzDrop (Fuzzy Droplet Predictor, https://fuzdrop.bio.unipd.it/predictor, accessed on 15 August 2022) [96].

All computer-generated structures of the RABV proteins analyzed in this study were generated using SWISS-MODEL [97] and ExPasy. The 3D structural models of human proteins that interact with the RABV proteins were generated by AlphaFold [98].

3. Results and Discussion

3.1. Predicted Disorder of the P-Protein and Its Suggested Functional Consequences

The P-protein of RABV, which is a 297-residue-long catalytic polymerase cofactor and regulatory protein that plays an important role in viral transcription and replication, was shown to display a significant amount of intrinsic disorder (see Figure 2). In fact, based on the PONDR® VSL2 outputs, roughly 49% of the protein residues were predicted to be intrinsically disordered (i.e., they have disorder scores exceeding the 0.5 threshold). Furthermore, almost 46% of its residues were predicted to be flexible (i.e., possessing disorder scores ranging from 0.15 to 0.5), indicating that 95% of this protein is expected to be either disordered of structurally flexible. Figure 2A also shows that the P-protein contains two long intrinsically disordered regions, IDD1 (residues 33–89) and IDD2 (residues 133–199), that flank an oligomerization domain, and a short disordered/flexible region (residues 242–252) located within the mostly ordered C-terminal domain (PCTD, residues 201–297). It was indicated that Negri bodies, also known as viral inclusion bodies in the host cytoplasm used for viral replication, are formed via the interaction of the P-protein oligomerization domain, IDD2, and the PCTD [99] with the intrinsically disordered regions of the N-protein, whereas the N-terminal part and IDD1 of the P-protein are dispensable [99,100].

High levels of intrinsic disorder in the RABV P-protein are further evidenced by the analysis of its X-ray crystal structure (PDB ID: 3OA1). In fact, although the 69–297 fragment of this protein was used in the crystallization experiments, the structure was eventually solved for less than half of this construct (residues 192–297), with the entire N-terminal half (residues 69–191) representing a region with missing electron density (i.e., highly flexible or disordered region). Furthermore, even within the solved structure of the C-terminal domain of the P-protein, some short regions were not modeled or incompletely modeled as well (residues 220–221, 231, 272–273, and 296–297) (PDB ID: 3OA1).

Importantly, as per the manually asserted information inferred from the sequence similarity and available in the UniProt database (https://www.uniprot.org/uniprotkb/P06747/entry, accessed on 15 August 2022), most of the IDD2 region of the RABV P-protein is expected to be engaged in binding to cytoplasmic dynein light-chains 1 and 2 (DYNLL1 and DYNLL2, see below). Furthermore, ANCHOR analysis [76,77] suggested that the P-protein contains three potential disorder-based binding sites, known as molecular recognition features (MoRFs), which are disordered regions that are expected to fold upon their interaction with specific partners, thereby driving protein–protein interactions. These are residues 34–90, 124–136, and 166–190. Therefore, it is likely that these segments (shown as gray-shaded areas in Figure 2A) might correspond to ligand binding sites of the RABV P-protein (see below).
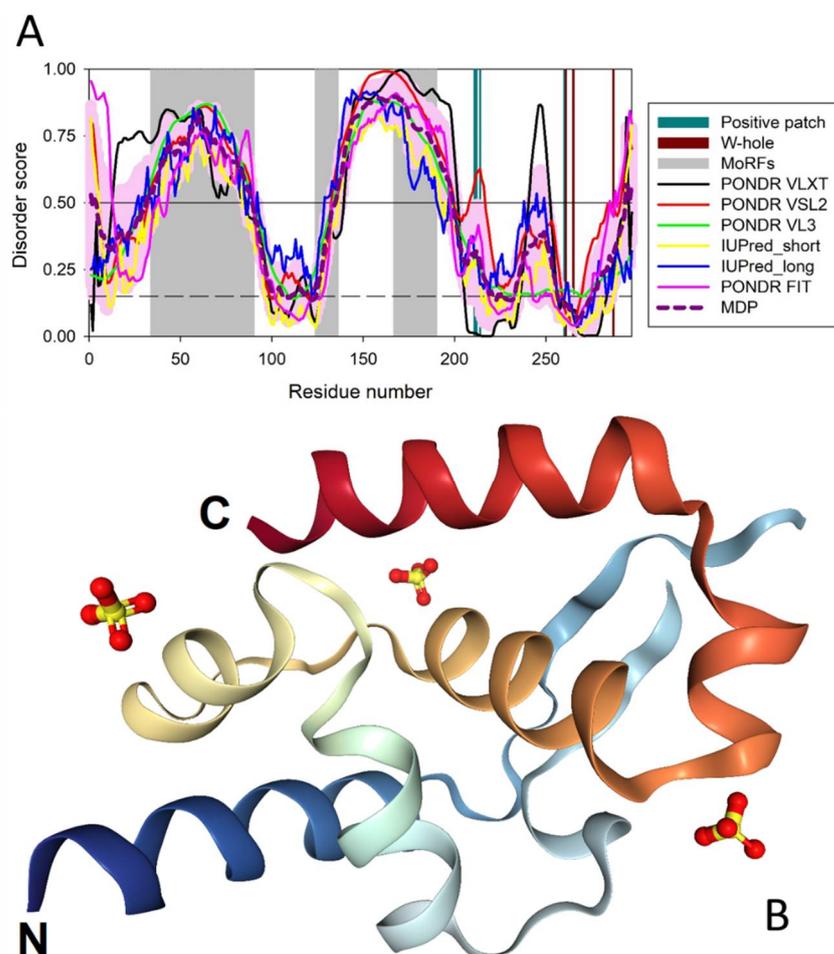
Figure 2. Structure and disorder in the RABV P-protein (UniProt ID: P06747). (**A**) Intrinsic disorder profile generated using data aggregated by the RIDAO platform. The profile also contains disorder/flexibility-related functional annotations and shows three predicted MoRFs (gray-shaded areas), positive patch residues (dark-cyan vertical bars), and W-hole residues (dark-red vertical bars). (**B**) Crystal structure of the C-terminal region of the RABV P-protein (residues 192–295) (PDB ID: 3OA1).

Additionally, the high disorder and structural flexibility content of the RABV P-protein suggests that the mechanism of action of this protein to suppress the type-I-interferon-mediated immune response may be based on the utilization of its disordered (or flexible) regions for interaction with the STAT proteins of cells. In fact, although the RABV P-protein binding site responsible for the STAT1/2 interaction is located within the ordered C-terminal domain (CTD, residues 186–297 [16,17,101–104]), the residues that made the greatest contribution to this interaction and that form the so-called W-hole (C261, W265, and M287) were predicted to be flexible by at least one of the disorder predictors used in this study, with the highest level of structural flexibility being expected for residue M287, which is 100% conserved among most lyssavirus P-proteins [16] and shows a mean disorder score of 0.31 ± 0.12 (ranging from 0.16 to 0.5, as per the outputs of IUPred_short and PONDR® VSL2, respectively (see Figure 2A). Similarly, the positive patch (residues K211, K212, K214, and R260), which is 100% conserved in the lyssavirus P-proteins and known to be responsible for interaction with the N-protein [16], is predicted to be flexible/disordered as well (see Figure 2A). This, again, suggests that structural flexibility plays a role in the interaction of the ordered CTD with partner proteins, including STAT1 and STAT2. It was pointed out that because of these C-terminal-domain-driven interactions of the RABV P-protein (CTD, residues 186–297 [101–104]) with host STAT proteins, the P-protein

represents the major interferon antagonist of the lyssavirus, thereby affecting the type-I-interferon (IFNα/β)-mediated innate immune response [16]. It was also pointed out that the interaction of the RABV P-protein with STATs is crucial for the development of the lethal rabies disease [16].

Another level of structural and functional complexity of this protein is given by the fact at it has multiple isoforms generated by the alternative initiation of the P-protein during viral transcription. In fact, alternative initiation generates isoforms P2, P3, P4, and P5, which differ from the canonical isoform P1 due to a lack of N-terminal residues 1–19, 1–52, 1–68, and 1–82. An obvious consequence of this truncation is the elimination of the first MoRF region of P1, suggesting that these isoforms might be characterized by different interactability. Curiously, although P3, P4, and P5 have all lost an N-terminal MoRF as expected, P2 was predicted to behave differently. In fact, despite missing the N-terminal residues predicted to be the MoRF in P1, this isoform gained three new N-terminal MoRFs (residues 1–14, 19–27, and 30–38).

The functional diversity of the RABV P-protein is further increased by the phosphorylation of its serine residues S63 and S64 by an unknown kinase (denoted rabies virus protein kinase (RVPK)) and residues S162, S210, and S271 by protein kinase C (PKC) [105], all located within IDPRs. A recent study revealed that the phosphorylation of the P3 isoform of the RABC P-protein at the S210 position resulted in a significant reduction in the nuclear localization modulated by the MT binding/bundling of P3 [106].

Therefore, due to the presence of regions with high intrinsic disorder content, several MoRFs, several phosphorylation sites, and the usage of alternative initiation, it is possible for the P-protein to serve many roles within the virus [10]. Furthermore, the P-protein isoforms were shown to differ in nucleocytoplasmic localization and microtubule (MT) association, mediated by several functional motifs, including the nuclear localization sequence (NLS, residues 211–214) and N- and C-terminally located nuclear export sequences (N-NES and C-NES, residues 49–58 and 223–232, respectively) [107–109]. For example, shorter isoforms (P3 to P5) lacking the N-terminally located NES are more nuclear and are capable of binding and bundling MTs [107]. As per the outputs of PONDR® VSL2, N-NES and NLS are located within IDPRs (residues 33–89 and 208–216, respectively; see Figure 2A).

Figure 2A shows that intrinsic disorder is unevenly distributed within the P-protein sequence. It is preferentially concentrated at its N-terminal and central regions (residues 1–200), with the C-terminal domain being predicted to possess a more ordered structure. In agreement with the results of the computational evaluation of the intrinsic disorder predisposition of this protein, a crystal structure was solved for residues 192–295 (see Figure 2B) (PDB ID: 3OA1). Curiously, as was already indicated, although a much longer fragment of the P-protein (residues 69–297) was used in the crystallization experiments, a very significant part of this polypeptide was not observed in the resulting structure, representing regions of missing electron density; i.e., regions with high conformational flexibility that preclude them from being crystallized.

Since the P-protein participates in the formation of Negri bodies [99] and can bind PML-bodies [110], we also checked the liquid–liquid phase separation (LLPS) potential of this protein by FuzDrop. This analysis revealed that the longest isoform (P1) of the P-protein is characterized by p_{LLPS} = 0.5276 and contains a droplet-promoting region (DPR, residues 134–184) located within the long central IDPR of this protein (residues 133–199). A recent systematic analysis revealed that this region indeed plays a crucial role in the formation of viral Negri body (NB)-like structures in infected cells [100]. Furthermore, it was shown that the deletion of residues 151–181 did not affect the ability of the RABV P-protein to form NB-like structures, indicating that only the amino-terminal part of IDD2 (residues 132–150) is required for this process [100].

Shorter isoforms of the P-protein, P2, P3, P4, and P5, showed p_{LLPS} values of 0.6561 (DPR, residues 115–166), 0.6832 (DPRs, residues 1–34 and 82–133), 0.4938 (DPRs, residues 1–19 and 66–115), and 0.3526 (DPR, residues 52–103), indicating that alternative initiation

dramatically affects the LLPS potential of the P-protein. According to the FuzDrop developers, proteins with $p_{LLPS} \geq 0.60$ are droplet-drivers, which can spontaneously undergo LLPS. Droplet-client proteins have pLLPS < 0.60 but possess droplet-promoting regions, which can induce their partitioning into condensates [96]. Therefore, according to the results of this analysis, P2 and P3 isoforms can potentially serve as droplet-drivers, whereas other isoforms (P1, P4, and P5) most likely represent droplet-clients.

FuzDrop also indicated that P1 contains nine regions with context-dependent interactions (residues 20–47, 52–64, 67–85, 94–101, 116–125, 129–138, 188–210, 252–257, and 273–282). These regions can potentially undergo disorder-to-order or disorder-to-disorder transitions but maintain conformational heterogeneity in the bound state and show sensitivity to the cellular context or post-translational modifications, potentially serving as regulatory engines of cellular pathways [111].

3.2. Disorder of the M-Protein and Its Suggested Functional Consequences

The M-protein is a 202-residue-long protein that plays a crucial role in the assembly and budding of the virion and engages in complete coverage of the ribonucleoprotein coil to keep it in a condensed bullet-shaped form (see Figure 1). It was found to be highly disordered as well, with 43% of this protein being composed of IDPRs (see Figure 3A). This suggests that the interactions of the M-protein with RelAp43 and other proteins in the NF-κB pathway [23] may induce the suppression of the antiviral response via the utilization of some advantages of intrinsic disorder.

This is in line with the results of several studies on the potential roles of intrinsic disorder in proteins that form the shells of several viruses (such as SARS-CoV-2, MERS-CoV, SARS-CoV, other CoVs, Nipah, Zika, HIV, and retroviruses) for viral transmission, immune evasion, and virulence [112–120]. This intrinsic disorder in the M-protein may also allow for the increased flexibility of the protein to aid in the regulation of virus budding. The M-protein has been shown to have the ability to create vesicles without any interaction with other viral proteins, suggesting that the flexibility of the protein allows it to induce budding by itself [3,13]. Figure 3A presents the disorder profile of the MATRX_RABVP protein and shows that a section in the middle of the protein, spanning roughly from residue 50 to residue 175, displays, on average, low disorder content and is likely to represent the structural domain of the protein, which, however, includes some disordered/flexible regions.

Curiously, region 115–151, which is essential for the glycoprotein (as per manually asserted information inferred from the sequence similarity and available in the UniProt database; see https://www.uniprot.org/uniprotkb/P08671/entry, accessed on 15 August 2022) binding, includes an IDPR (residues 129–141, as per PONDR® VLXT, or residues 130–137, as per PONDR® VSL2), indicating that the intrinsic disorder (or structural flexibility) of this region can contribute to its interactability. Furthermore, the M-protein contains a PPxY motif (residues 35–38), which is commonly found in viral proteins capable of manipulating the autophagic machinery to prevent the autophagic degradation of viruses [121]. This PPxY motif is included in the long N-terminal IDPR (residues 1–48) (see Figure 3A).
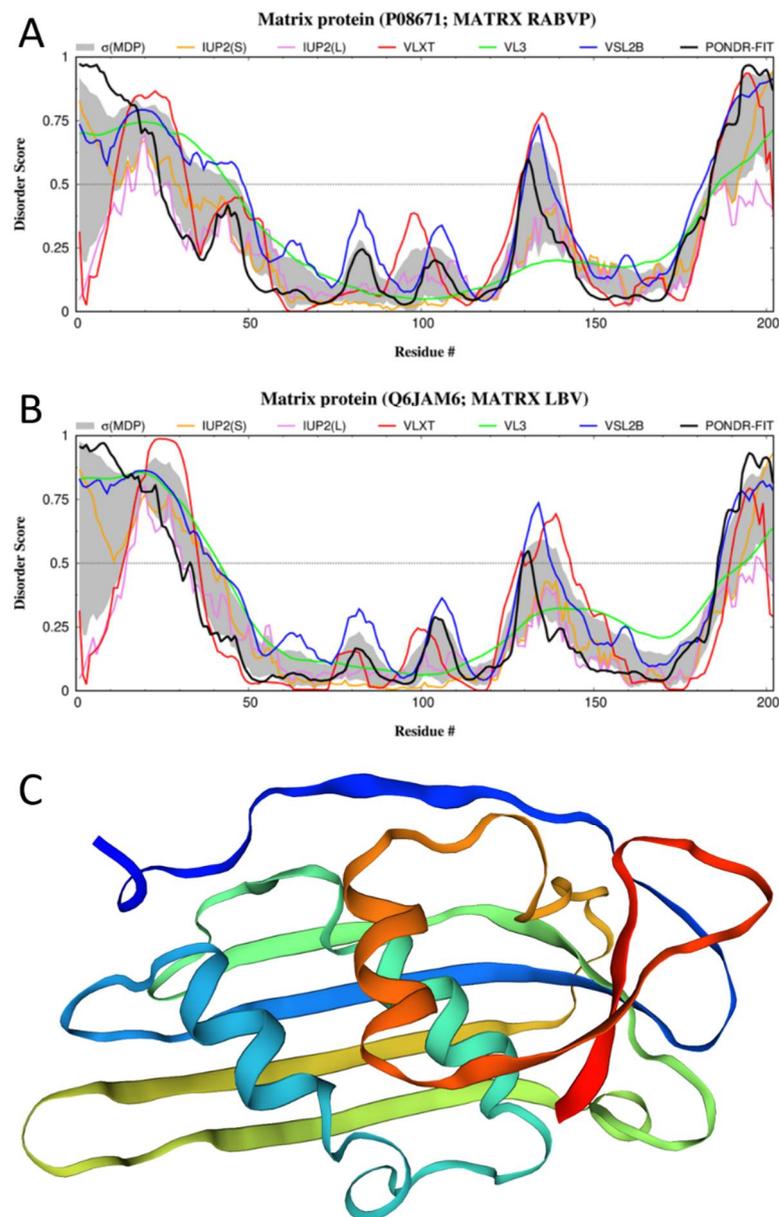
Figure 3. Structure and disorder in the M-protein from RABV (UniProt ID: P08671). (**A**) Intrinsic disorder profile (disorder score vs. residues number (residues #)) generated for the M-protein from the RABV strain VP by RIDAO. (**B**) Intrinsic disorder profile generated for the Lagos bat virus matrix protein (UniProt ID: Q6JAM6) by RIDAO. (**C**) A structural model for the 30–202 fragment of the RABV M-protein that was created by SWISS-MODEL [97] using the structure of the Lagos bat virus matrix protein (PDB ID: 2W2S; [122]) as a template, which shows sequence identity to the query M-protein of 76.73%.

Although no structural information is currently available for the M-protein of RABV, we used SWISS-MODEL (https://swissmodel.expasy.org/, accessed on 15 August 2022) to create homology models for this protein. Figure 3C shows a model for the 30–202 fragment of the RABV M-protein that was created using the structure of the Lagos bat virus matrix protein (PDB ID: 2W2S; [122]; UniProt ID: Q6JAM6) as a template with sequence identity to the query M-protein of 76.73%. A comparison of Figure 3A,B illustrates the remarkable similarity between the per-residue intrinsic disorder predispositions of the Lagos bat virus matrix protein and the RABV M-protein, thereby providing the intrinsic-disorder-based validation of the selection of the Lagos bat virus matrix protein as a template.

Finally, FuzDrop analysis indicated that although the M-protein shows a low probability of spontaneous LLPS (p_{LLPS} = 0.2220), this protein contains one C-terminally located DPR (residues 199–202), which is included in the IDPR (residues 182–202), indicating that the M-protein can act as the droplet-client. FuzDrop also identified regions 16–29, 121–131, and 172–190 as regions with context-dependent interactions [111].

3.3. Disorder of the N-Protein and Its Suggested Functional Consequences

The 450-residue-long N-protein is a viral RNA-binding protein that encapsulates the genome in a ratio of one N-protein per nine ribonucleotides. The long C-terminal IDPR and several shorter IDPRs make up roughly 31% of the protein (see Figure 4A). The N-protein, which is the most transcriptionally abundant protein during infection [123], has been shown to encapsulate the viral genomic RNA to protect it from nucleases and form a complex with the P-protein during replication [124]. The RABV P-protein binds to the RNA-free N°-protein through the N-terminus [100,125], which is predicted to be highly intrinsically disordered (see Figure 2A), whereas the N-terminal of the P-protein CTD is responsible for its interaction with the RNA-bound N-protein (see below).
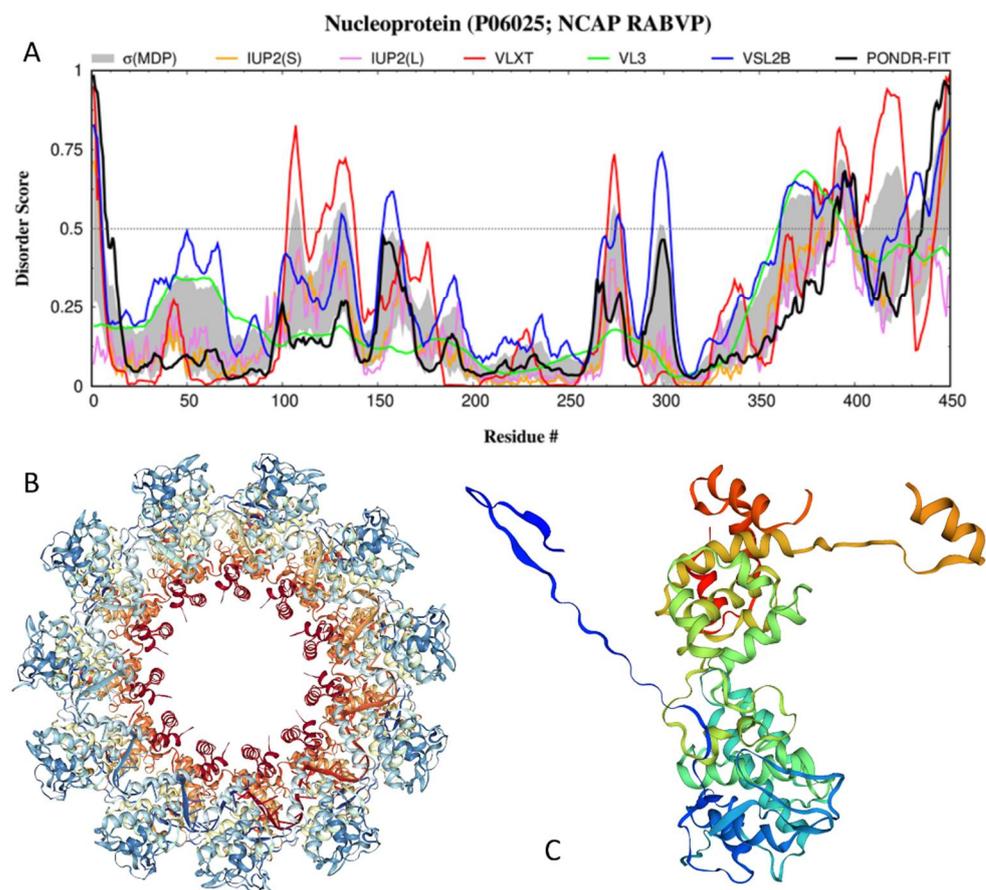

Figure 4. Structure and disorder in the N-protein from the RABV strain PV (UniProt ID: P06025). (**A**) Intrinsic disorder profile generated for the M-protein from the RABV strain VP by RIDAO. (**B**) Crystal structure of the N-protein from the RABV strain ERA (which is 99.11% identical to the N-protein from the RABV strain PV) in complex with RNA (PDB ID: 2GTT; [126]), where the protomers of the N-protein are organized in an undecameric ring. (**C**) Crystal structure of the N-protein protomer computationally taken out of the undecameric homo-oligomer and demonstrating the presence of two "arms" in the structure (residues 6–28 and 349–414).

The complex formed during replication, called a Negri body, is an inclusion in the host's cytoplasm that is formed from interactions of the highly disordered C-terminal

region of the N-protein with the intrinsically disordered regions of the P-protein [127]. The complex prevents non-specific RNA binding and the phosphorylation of the RNA [6]. The N-protein is predicted to have one MoRF (residues 406–411) located within the long C-terminal IDPR, which also includes a phosphoserine at position 389. A large region in the middle of the protein with little disorder suggests that the protein serves a largely structural purpose. The N-protein also functions to prevent the activation of the antiviral innate immune response receptor RIG-I-mediated antiviral response [19,20]. Although the actual information on the molecular mechanisms of this suppression is currently unavailable, it is tempting to hypothesize that the MoRF-containing IDPR found in the N-protein may serve to aid in suppressing the antiviral response. Figure 4A shows the results of the disorder prediction for the NCAP_RABVP protein. The crystal structure of the N-protein from the RABV strain ERA (which is 99.11% identical to the N-protein from the RABV strain PV) in complex with RNA was solved (PDB ID: 2GTT; [126]). Figure 4B shows that in this RABV nucleoprotein–RNA complex, the N-proteins are organized in an undecameric ring. In such a complex, the two core domains of the nucleoprotein clamp around the RNA at their interface and shield it from the environment [126]. Polymerization of the nucleoprotein is achieved by domain exchange between protomers, with flexible hinges allowing nucleocapsid formation. It is likely that the high structural flexibility and pliability of the C-terminal region of the protein, which is predicted to be highly intrinsically disordered, make it able to more tightly cover bound RNA. The nucleoprotein and the RNP core are able to adopt different conformations as a result at different periods of the viral cycle [127]. This important observation is illustrated in Figure 4C, which shows the structure of the N-protein protomer and demonstrates the presence of two "arms" in the structure (residues 6–28 and 349–414). Importantly, the C-terminal arm contains a MoRF. In addition to these functional arms, each protomer has several regions of missing electron density (residues 1–5, 104–117, 186–188, 374–397, and 449–450). Furthermore, according to the FuzDrop analysis, one can find four regions in the N-protein with context-dependent interactions (residues 105–115, 284–318, 367–396, and 398–411), which overlap, include, or are included in disordered/flexible regions of this protein (residues 103–111, 294–303, 361–400, and 403–428) (see Figure 4A).

Although, based on the FuzDrop analysis, the N-protein is expected to have a low probability of spontaneous LLPS ($p_{LLPS} = 0.1405$) and does not include any DPRs, this RNA-binding protein is invariantly present in Negri bodies (NBs) [100]. It is likely that the involvement of the N-protein in NB biogenesis is linked to the ability of this protein to bind both viral RNA and the P-protein. In fact, no NBs were found when limiting concentrations of one of these proteins were expressed in model experiments [100]. Furthermore, even when the RABV P- and N-proteins were expressed alone (i.e., without viral RNA), they were capable of forming NB-like structures [100]. These observations indicated that in the resulting N-P inclusions, the RABV N-protein is likely bound to cellular RNAs, forming N-RNA complexes similar to viral nucleocapsids [125]. It is known that different regions of the P-protein are utilized in the interaction with the N-protein, where the disordered N-terminal domain interacts with the RNA-free N°-protein, whereas the P-protein CTD binds to the RNA-associated N-protein [100,125]. As was already indicated, the positive patch within the N-terminal region of the P-protein CTD that is actually responsible for the interaction with the RNA-associated N-protein contains flexible/disordered residues, further emphasizing the potential role of structural disorder/flexibility in NB biogenesis.

3.4. Disorder of the G-Protein and Its Suggested Functional Consequences

The G-protein is a 524-residue-long type I transmembrane protein with a long extravirion region (residues 20–459), a transmembrane helix (residues 460–480), and an intravirion domain (residues 481–524). Being located on the surface of RABV particles, the G-protein controls receptor binding and the release of the viral ribonucleoprotein (RNP) in the cytoplasm via pH-dependent membrane fusion, thereby playing a crucial role in the cell entry and in vivo spread [128]. Furthermore, it was shown that the G-protein (in

particular, its ectodomain) accumulates adaptive mutations that improve the release of infectious viral particles [129].

The G-protein is synthesized as a precursor with the N-terminal signal peptide (residues 1–19), which is removed during the maturation of this protein. Similar to the proteins found in the envelopes of other viruses, the RABV G-protein forms homotrimers on the surface of the virion that are responsible for the attachment of the virus to the host cellular receptors, such as the muscular form of the nicotinic acetylcholine receptor (nAChR), the neuronal cell adhesion molecule (NCAM), and the p75 neurotrophin receptor (p75NTR). There are approximately 400 such trimeric spikes, which are tightly arranged on the surface of the virus. The C-terminal domain of the G-protein (residues 258–505) is essential for trimer stability [8]. Figure 5A shows that the G-protein is mostly ordered and contains relatively few IDPRs, which is a characteristic feature of spike/glycoproteins of many other viruses. The most promising predicted disordered region is the C-terminal IDPR (residues 486–524; see Figure 5A), which corresponds to the intravirion domain of this protein (https://www.uniprot.org/uniprotkb/P08667/entry, accessed on 15 August 2022) that is engaged in the interaction with the matrix protein [13]. Several flexible regions located within the extravirion domain serve as aids in the interaction of the G-protein with surface molecules of the host cell [24]. There are three glycosylation cites in this protein (asparagine residues 56, 266, and 338) and a C-terminally located lipidation site, S-palmitoyl cysteine 479.

There is currently no structural information for the G-protein from the RABV strain PV. Therefore, we used SWISS-MODEL to create a homology model for this protein. Figure 5C shows a model for the 20–424 fragment of this protein that was generated using the known structure of the G-protein from the RABV strain CVS-11 (PDB ID: 6LGW [130]; UniProt ID: O92284) as a template, with sequence identity to the query G-protein of 91.48%. This structure is characterized by a highly elongated form and the presence of a C-terminal "arm" (residues 400–416). In the original structure of the G-protein from the RABV strain CVS-11, there are several regions of missing electron density (residues 21–24, 95–101, 133–143, 202–204, and 414–429).

Furthermore, the authors of this study present a structure for another form of this protein (PDB ID: 6LGX), which shows a different pattern of missing electron density regions (residues 21–25, 91–105, 131–147, 274–294, and 427–463). Since these two structures were resolved under different conditions (at ~pH-6.5 in the complex with a neutralizing antibody 523–11 (PDB ID: 6LGW) and at ~pH-8.0 in free form (PDB ID: 6LGX)), these observations suggest that this G-protein structure is characterized by its noticeable sensitivity to environmental conditions. The authors pointed out that the basic-to-acidic pH change results in large re-orientations of the three domains found in the G-protein from the RABV strain CVS-11, leading to concomitant domain-linker reconstructions that switch from a bent hairpin conformation into an extended conformation [130]. These low-pH-induced structural transitions within the domain-linker region are related to the functionality of the G-protein, playing important roles in G-protein-mediated membrane fusion [130]. Figure 5A,B show that the G-proteins from the RABV strains PV and CVS-11 are characterized by very similar disorder profiles, suggesting that the observations made for the structural peculiarities of the RABV G-protein from the CVS-11 strain are applicable to the G-protein from the RABV strain PV as well.

Similar to the N-protein, the glycoprotein of RABV is characterized by a low probability of spontaneous LLPS ($p_{LLPS} = 0.1351$) and does not include any DPRs but contains five regions with context-dependent interactions (residues 18–29, 218–223, 229–239, 434–441, and 502–508), which overlap, include, or are included in the disordered/flexible regions of this protein (residues 214–225, 422–442, and 486–524) (see Figure 5A).
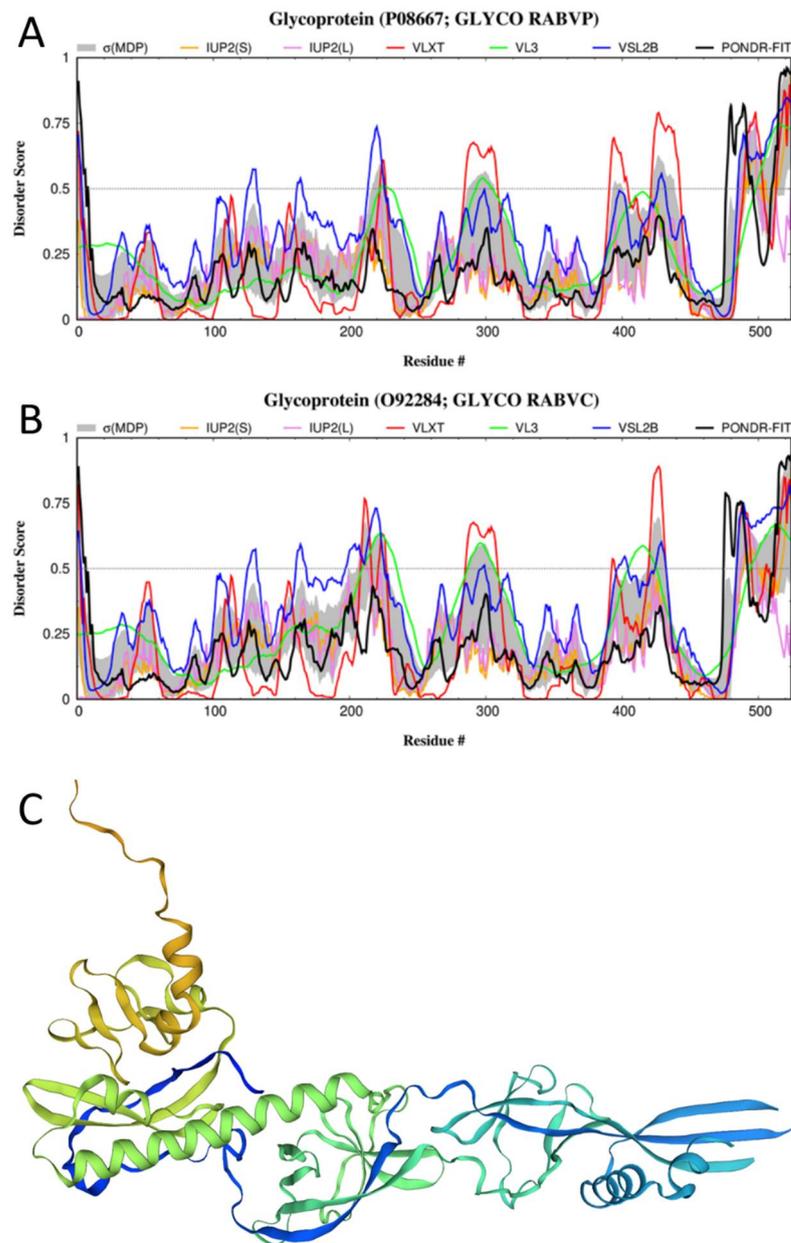
Figure 5. Structure and disorder in the G-protein from RABV. (**A**) Intrinsic disorder profile generated for the G-protein from the RABV strain VP (UniProt ID: P08667) by RIDAO. (**B**) Intrinsic disorder profile generated for the G-protein from the RABV strain CVS-11 by RIDAO (UniProt ID: O92284). (**C**) A structural model for the 20–424 fragment of the G-protein from the RABV strain VP generated by SWISS-MODEL [97] using the known structure of the G-protein from the RABV strain CVS-11 (PDB ID: 6LGW [130]; UniProt ID: O92284) as a template, with sequence identity to the query G-protein of 91.48%.

3.5. Disorder of the L-Protein and Its Suggested Functional Consequences

With an amino acid sequence of 2124 residues, the L-protein is the longest protein in the RABV proteome. This protein is an RNA-directed RNA polymerase (RdRp) that catalyzes the transcription of viral mRNAs as well as their polyadenylation and capping. It has several functional regions, such as an RdRp catalytic domain (residues 611–799), a mononegavirus-type SAM-dependent 2'-O-MTase domain (residues 1674–1871) included in the C-terminal region that is involved in the interaction with the P-protein (residues 1562–2127) and contains a disorder-based interaction site, and a MoRF (residues 1631–1638). In addition

to interactions with the P-protein, the L-protein may form homodimers. Although the L-protein includes many disordered or flexible regions, its overall intrinsic disorder level is relatively low (see Figure 6A).

It is possible that the need for interplay between ordered and disordered features in this protein reflects its purpose to serve as a low-fidelity viral polymerase, as intrinsic disorder and structural flexibility 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]. The polymerase activity of the L-protein depends on the identity of residues upstream of the protein, as well as the identity of C-terminal residues [131]. Figure 6C shows the structural model for the L-protein from the RABV strain PV built using the cryo-EM structure of the large structural protein from the RABV strain SAD B19 (sequence identity: 98.68%) complexed with a fragment of the P-protein (PDB ID: 6UEB, [132]; UniProt ID: P16289) as a template. Due to their high sequence similarity, the disorder profiles of the L-proteins from RABV strains SAD B19 and PV are almost identical (cf. Figure 6A,B). Since in the aforementioned cryo-EM structure of the L-P complex, residues 1–27 of the L-protein from the RABV strain SAD B19 constitute a region of missing electron density, it is likely that this N-terminal region is disordered in the L-protein from the RABV strain PV as well.
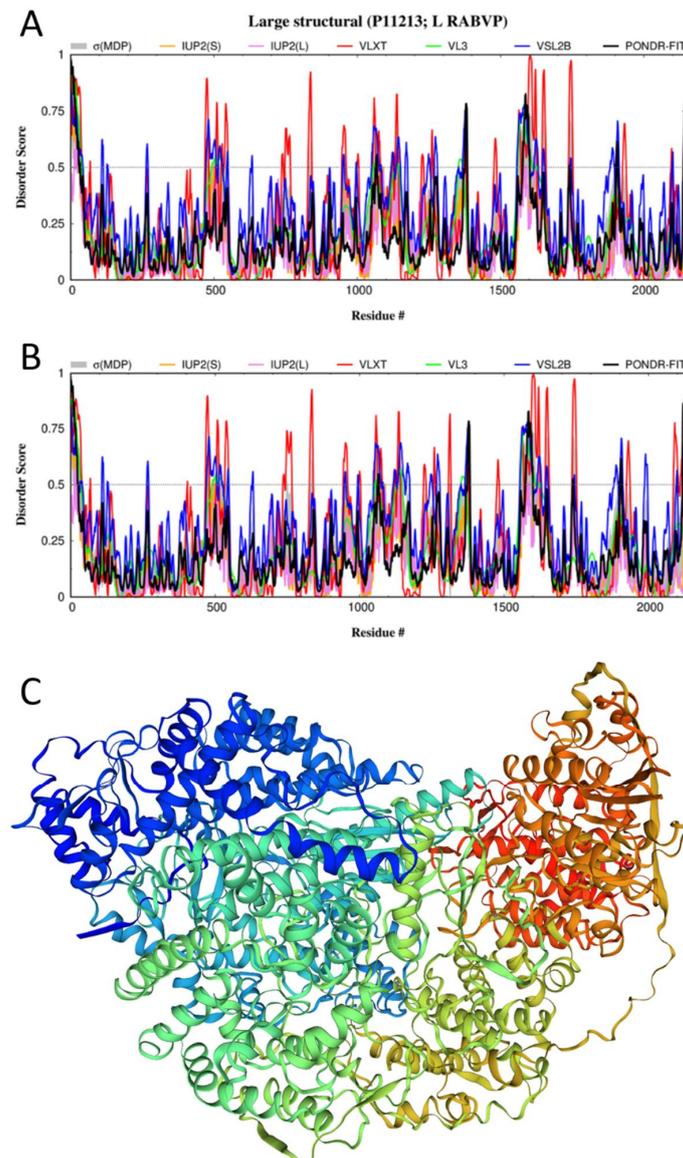

Figure 6. Structure and disorder in the L-protein from RABV. (**A**) Intrinsic disorder profile generated

for the L-protein from the RABV strain VP (UniProt ID: P11213) by RIDAO. (**B**) Intrinsic disorder profile generated for the L-protein from the RABV strain SAD B19 by RIDAO (UniProt ID: P16289). (**C**) A structural model for the L-protein from the RABV strain PV built by SWISS-MODEL [97] using the cryo-EM structure of the large structural protein from the RABV strain SAD B19 (sequence identity: 98.68%) complexed with the fragment of the P-protein (PDB ID: 6UEB, [132]; UniProt ID: P16289) as a template.

FuzDrop indicated that the L-protein has a low p_{LLPS} value of 0.11860 and does not contain any DPRs. However, this protein possesses 24 regions with context-dependent interactions (residues 5–12, 104–109, 416–434, 456–480, 507–514, 566–571, 672–677, 821–833, 852–859, 878–899, 903–911, 924–929, 982–987, 1014–1022, 1152–1157, 1232–1238, 1383–1388, 1390–1399, 1562–1570, 1655–1667, 1731–1753, 1980–1985, 2086–2092, and 2127–2132). Many of these regions are related to the IDPRs identified in the L-protein (e.g., residues 1–42, 108–114, 466–480, 504–512, 827–839, 1555–1599, 1642–1656, 1735–1750, 2093–2096, and 2137–2142).

3.6. Intrinsic Disorder in Human Proteins Interacting with the RABV Proteins

3.6.1. Host Interactors of the P-Protein

In addition to being involved in interactions with the RABV L- and N-proteins and viral ribonucleocapsids, the P-protein is known to bind to a multitude of host proteins, such as dynein light-chain 1 and 2 (DYNLL1 (UniProt ID: P63167) and DYNLL2 (UniProt ID: Q96FJ2), respectively) [133], as well as host signal transducer and activator of transcription 1-alpha/beta and signal transducer and activator of transcription 2 (STAT1 (UniProt ID: P42224) and STAT2 proteins (UniProt ID: P52630), respectively) [16,17,101–104], promyelocytic leukemia (PML) protein (UniProt ID: P29590) [110], the ribosomal protein L9 (UniProt ID: Q02878) [15], STAT3 (UniProt ID: P40763), nucleolin (NCL; UniProt ID: P19338), focal adhesion kinase (FAK; UniProt ID: Q05397), Janus kinase 1 (JAK1; UniProt ID: P23458), inhibitor of nuclear factor kappa-B kinase subunit epsilon (IKKε; UniProt ID: Q14164), Beclin-1 (BECN1; UniProt ID: Q14457), tubulin alpha (TUB-α; UniProt ID: Q71U36 for tubulin alpha-1A chain), tubulin beta (TUB-β; UniProt ID: Q9H4B7 for tubulin beta-1 chain), ATP-binding cassette sub-family E member 1 (ABCE1; UniProt ID: P61221), T-complex protein 1 subunit gamma (CCTγ; UniProt ID: P49368), Hsp90 co-chaperone Cdc37 (CDC37; UniProt ID: Q16543), and heat shock protein HSP 90-alpha (Hsp90AA1; UniProt ID: P07900) [71]. Furthermore, the P-protein can interact with complex I in mitochondria, leading to mitochondrial dysfunction, the increased generation of reactive oxygen species (ROS), and oxidative stress [134]. Being the largest and most complicated component of the respiratory chain, complex I contains 45 subunits [135]. Unfortunately, the exact targets of the P-protein within this complex are unknown. Therefore, the proteins forming complex I were not included in the analysis.

3.6.2. Host Interactors of the M-Protein

Several human proteins that interact with the RABV M-protein were established. The list includes RelAp43 (which is a splicing variant of RelA (UniProt ID: Q04206) and other proteins in the NF-κB pathway [23], as well as V-type proton ATPase catalytic subunit A (ATP6V1A; UniProt ID: P38606), E3 ubiquitin-protein ligase NEDD4 (NEDD4; UniProt ID: P46934), transcriptional coactivator YAP1 (UniProt ID; P46937), eukaryotic translation initiation factor 3 subunit H (EIF3H, UniProt ID: O15372), JAK1 (UniProt ID: P23458), and STAT1 (UniProt ID: P63167) [71]. Note that the M-protein shares two human interactors (JAK1 and STAT1) with the P-protein [71].

3.6.3. Host Interactors of the N-Protein

Very few host proteins were shown to act as physical partners of the RABV N-protein. These include heat shock 70 kDa proteins 1A (HSPA1A, UniProt ID: P0DMV8) and 1B

(HSPA1B, UniProt ID: P0DMV9), prefoldin subunit 1 (PFDN1, UniProt ID: O60925), and CCTγ (UniProt ID: P49368), with CCTγ being a shared partner for the RABV P- and N-proteins [71].

3.6.4. Host Interactors of the G-Protein

The major biological function of the G-protein is cell entry via interaction with RABV entry receptors, such as nicotinic acetylcholine receptors (nAChR), the neuronal cell adhesion molecule (NCAM1; UniProt ID: P13591), the low-affinity p75 neurotrophin receptor (p75NTR, also known as tumor necrosis factor receptor superfamily member 16; UniProt ID: P08138), and metabotropic glutamate receptor subtype 2 (mGluR2; UniProt ID: Q14416) [136–139]. nAChRs are assemblies of five subunits, which are arranged symmetrically around a central pore. It is known that the neuronal subtypes of nAChRs (which serve as the receptors of the RABV G-protein) exist as various homomeric or heteromeric (at least one α and one β) combinations of twelve different nicotinic receptor subunits: $\alpha_2-\alpha_{10}$ and $\beta_2-\beta_4$ (with some of the neuronal nAChR subtypes being $(\alpha_4)_3(\beta_2)_2$, $(\alpha_4)_2(\beta_2)_3$, $(\alpha_3)_2(\beta_4)_3$, $\alpha_4\alpha_6\beta_3(\beta_2)_2$, and $(\alpha_7)_5$). Therefore, there are multiple possibilities for the binding of the G-protein to nAChR. Physical interactions of nicotinic acetylcholine receptor alpha 1 (nAChR α1 or CHRNA1; UniProt ID: P02708) [140] and nAChr α7 (CHRNA7, UniProt ID: P36544) [141] with the RABV G-protein were demonstrated. Furthermore, the G-protein can bind host microtubule-associated serine/threonine-protein kinases 1 and 2 (MAST1 (UniProt ID: Q9Y2H9) and MAST2 (UniProt ID: Q6P0Q8), respectively), tyrosine-protein phosphatase non-receptor type 4 (PTPN4; UniProt ID: P29074), disks large homolog 2 (DLG2; UniProt ID: Q15700), multiple PDZ domain protein (MPDZ; UniProt ID: O75970), and synaptosomal-associated protein 25 (SNAP25; UniProt ID: P60880) [71].

3.6.5. Host Interactors of the L-Protein

The only human partner of the L-protein is DYNLL1, which is shared with the P-protein [71].

3.6.6. Prevalence of Intrinsic Disorder in Human Proteins Interacting with RABV

Detailed characterizations of the prevalence of functional disorder in each of the human proteins interacting with the RABV P-, M-, N-, G-, and L-proteins are systemized in Supplementary Figures S1–S5, respectively. The corresponding analyses were conducted using a set of computational tools, such as RIDAO, STRING, D^2P^2, and AlphaFold. This revealed a very high level of disorder in the majority of the proteins from this dataset, with the entire set being characterized by a 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 [142,143]).

Supplementary Figures S1–S5 show that all of these proteins contain multiple IDPRs of various lengths. Many proteins contain multiple MoRFs, and almost all human proteins in this dataset are densely decorated by a multitude of different PTMs. These observations suggest that intrinsic disorder in this these proteins might be related to their functionality, playing a role in their binding promiscuity, as evidenced by dense PPI networks centered on these proteins.

These observations are further illustrated in Figure 7, which presents the global intrinsic disorder characteristics of 37 human proteins. In fact, Figure 7A shows there is no 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%). This classification is accepted in field practice to group proteins based on their PPIDR values [87], where proteins with PPIDR < 10% are considered ordered or mostly ordered; proteins with 10% ≤ PPIDR < 30% are considered moderately disordered; and proteins with PPIDR ≥ 30% are considered highly disordered [87]. Furthermore, PPIDR values for 13 proteins (NCL (P19338; 86.20%), YAP1 (P46937; 82.94%), SNAP25 (P60880; 78.16%), MAST2 (Q6P0Q8; 73.75%), MAST1 (Q9Y2H9; 68.60%), CDC37 (Q16543; 65.08%), RelAp43

(Q04206; 64.61%), NEDD4 (P46934; 62.17%), L9 (Q02878; 61.46%), BECN1 (Q14457; 58.00%), p75NTR (P08138; 57.14%), PML (P29590; 54.06%), and PFDN1 (O60925; 50%)) exceed 50%, indicating that 35.1% of human proteins interacting with RABV are expected to be extremely disordered. These levels of disorder in human RABV interactors are comparable to those observed in the entire human proteome, where out of 20,317 proteins, 12,363 proteins (60.8%) and 7590 (37.3%) are characterized by PPIDR $\geq$ 30% and PPIDR $\geq$ 50%, respectively.

Figure 7B shows the ΔCH-ΔCDF plot (a combined binary predictor of intrinsic disorder) that verifies a global prevalence of intrinsic disorder in 37 human proteins interacting with the RABV proteins. The ΔCH-ΔCDF plot provides the means for the evaluation of the flavors of intrinsic disorder. Figure 7B shows that quadrant Q1 (bottom right corner) contains 22 proteins that are predicted to be ordered by both predictors; quadrant Q2 (bottom left corner) includes 10 proteins, which are predicted to be ordered/compact by the 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); and quadrant Q3 (top left corner) includes 4 highly disordered regions (native coils or native pre-molten globules), which are predicted to be disordered by both predictors. Finally, one protein in quadrant Q4 (top right corner) is predicted to be disordered by the CH-plot and ordered by the CDF-plot.

Therefore, 15 human proteins that interact with the RABV proteins are predicted to contain very noticeable levels of disorder (i.e., they are located outside quadrant Q1). This correlates well with the results shown in Figure 7A, where 13 proteins are located within the dark-red segment.

The apparent discrepancies between the data shown in Figure 7A,B are rooted in the principle differences in the tools utilized for these analyses, where Figure 7A represents the outputs of the per-residue predictor, whereas Figure 7B reports data generated by the so-called binary predictors; i.e., tools that classify query proteins as mostly ordered or mostly disordered. Obviously, mostly ordered proteins might contain noticeable levels of disordered residues, whereas mostly disordered proteins might possess noticeable levels of ordered residues.

Not surprisingly, all 37 proteins in the analyzed set were found to form a rather dense PPI network (see Figure 8A), in which, on average, each protein interacts with at least 11 partners.

In this intraset PPI network, the five most significantly enriched biological processes were viral process (GO:0016032; $p = 3.49 \times 10^{-8}$), immune effector process (GO:0002252; $p = 0.00030$), interspecies interaction between organisms (GO:0044419; $p = 0.00030$), intracellular transport (GO:0046907; $p = 0.00057$), and cellular component organization (GO:0016043; $p = 0.00057$). Among the molecular functions, the five most significantly enriched were ubiquitin-like protein ligase binding (GO:0044389; $p = 7.02 \times 10^{-7}$), protein binding (GO:0005515; $p = 1.22 \times 10^{-6}$), ubiquitin protein ligase binding (GO:0031625; $p = 6.83 \times 10^{-5}$), enzyme binding (GO:0019899; $p = 8.27 \times 10^{-5}$), and binding (GO:0005488; $p = 8.27 \times 10^{-5}$). Finally, the five most significantly enriched cellular components were postsynapse (GO:0098794; $p = 1.79 \times 10^{-6}$), cytosol (GO:0005829; $p = 1.79 \times 10^{-6}$), synapse (GO:0045202; $p = 2.96 \times 10^{-5}$), cell junction (GO:0030054; $p = 0.00010$), and neuron projection (GO:0043005; $p = 0.00017$).
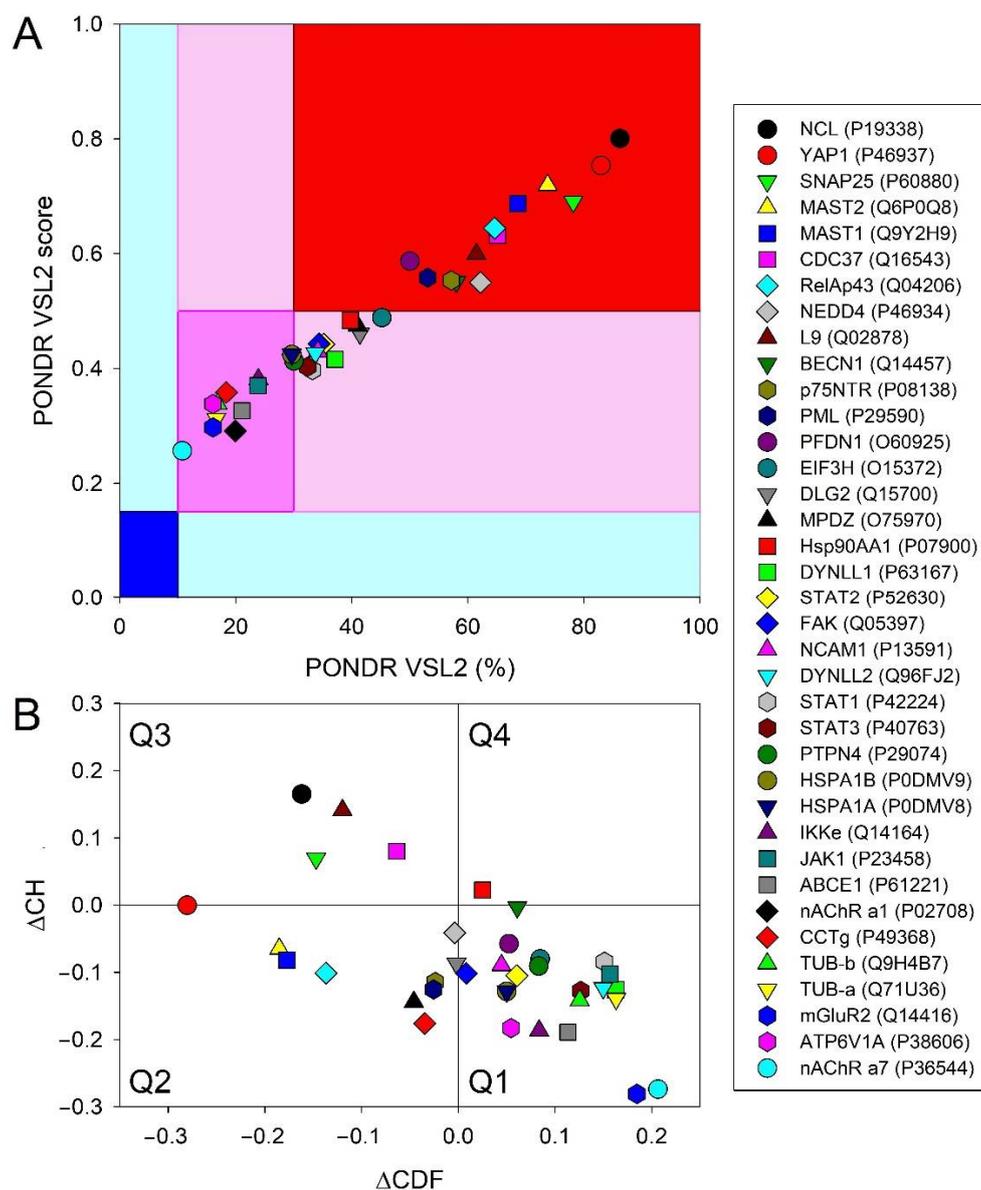
Figure 7. Evaluation of global disorder in 37 human proteins interacting with the RABV proteins. (**A**) PONDR® VSL2 output for 37 human proteins. The PONDR® VSL2 score is the average disorder score (ADS) for a protein. PONDR® VSL2 (%) is the percent of predicted disordered residues (PPDR). i.e., residues with disorder scores above 0.5. Color blocks indicate regions in which proteins are mostly ordered (blue and light blue), moderately disordered (pink and light pink), or mostly disordered (red). If the two parameters agree, the corresponding part of the background is dark (blue or pink), whereas light blue and light pink reflect areas in which only one of these criteria applies. (**B**) Charge-hydropathy and cumulative distribution function (CH-CDF) plot. The Y-coordinate is calculated as the distance of the corresponding protein from the boundary in the CH plot. The X-coordinate is calculated as the average distance of the corresponding protein's CDF curve from the CDF boundary. The quadrant in which the protein is located determines its classification. Q1, protein predicted to be ordered by the CH-plot and CDF. Q2, protein predicted to be ordered by the CH-plot and disordered by the CDF-plot. Q3, protein predicted to be disordered by the CH-plot and CDF. Q4, protein predicted to be disordered by the CH-plot and ordered by CDF.
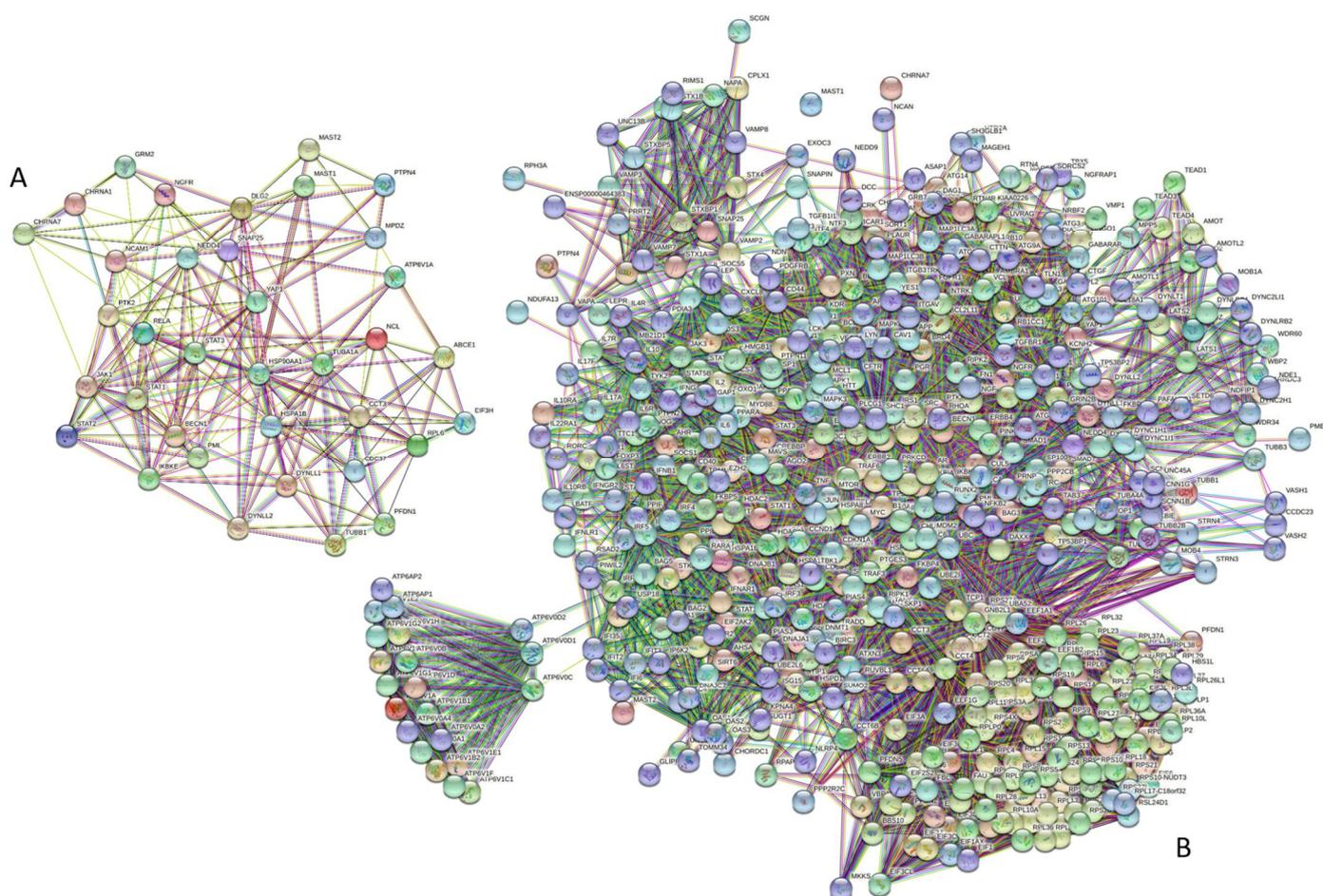
Figure 8. Intraset and set-based interactivity of human proteins engaged in interactions with the RABV proteins. (**A**) STRING-generated PPI network with the analyzed set of human proteins. To include all proteins in the network, a low confidence of 0.15 was used as the minimum required interaction score in this case. This network includes 36 proteins linked by 204 interactions. The resulting average node degree of this network is 11.3, and its average local clustering coefficient (which defines how close its neighbors are to being a complete clique; the local clustering coefficient is equal to 1 if every neighbor connected to a given node Ni is also connected to every other node within the neighborhood, and it is equal to 0 if no node that is connected to a given node Ni connects to any other node that is connected to Ni) is 0.578. Since the expected number of edges in a network of the same size for proteins randomly selected from the human proteome is 128, this network is characterized by a PPI enrichment *p*-value of 4.7×10^{-10}. (**B**) The STRING-generated PPI network centered on human proteins interacting with the RABV proteins. Note that the number of interactors in STRING is limited to 500. This network, generated with a high confidence score of 0.7, includes 536 proteins connected by 11,358 interactions. The average node degree and average local clustering coefficient of this PPI are 42.4 and 0.631, respectively; its PPI enrichment *p*-value is $<10^{-16}$.

We also analyzed the set-based interactivity of these human proteins. To this end, STRING was used to generate a PPI network that includes first-shell interactors (see Figure 8B). The resulting highly interlinked interactome includes 536 proteins connected by 11,358 interactions. Therefore, this interactome is characterized by an average node degree of 42.4, and it shows an average local clustering coefficient of 0.631. The expected number of interactions for the set of proteins of its size is 5.089, indicating that this PPI network, centered on human proteins interacting with the RABV proteins, has significantly more interactions than expected (PPI enrichment *p*-value is $< 10^{-16}$). Looking at the functional enrichment of this network based on Gene Ontology (GO) terms revealed that the five

most significantly enriched biological processes were interspecies interaction between organisms (GO:0044419; $p = 6.95 \times 10^{-102}$), viral process (GO:0016032; $p = 1.60 \times 10^{-99}$), symbiotic process (GO:0044403; $p = 1.60 \times 10^{-99}$), translational initiation (GO:0006413; $p = 7.45 \times 10^{-88}$), and cellular response to organic substance (GO:0071310; $p = 6.37 \times 10^{-76}$). The five most significantly enriched functions were structural constituent of ribosome (GO:0003735; $p = 5.85 \times 10^{-66}$), protein binding (GO:0005515; $p = 3.38 \times 10^{-57}$), binding (GO:0005488; $p = 2.02 \times 10^{-52}$), enzyme binding (GO:0019899; $p = 3.08 \times 10^{-46}$), and RNA binding (GO:0003723; $p = 2.09 \times 10^{-37}$). Among the cellular components, the five most significantly enriched were cytosol (GO:0005829; $p = 6.45 \times 10^{-106}$), protein-containing complex (GO:0032991; $p = 2.61 \times 10^{-82}$), cytosolic ribosome (GO:0022626; $p = 5.33 \times 10^{-81}$), cytoplasm (GO:0005737; $p = 6.02 \times 10^{-68}$), and ribosomal subunit (GO:0044391; $p = 5.65 \times 10^{-65}$).

Although a detailed description of the potential roles of intrinsic disorder in the functionality of human proteins shown to interact with the RABV proteins is outside the scope of this study, it is clear that all of the proteins analyzed here (i.e., RABV N-, L-, P-, M-, and G-proteins and 37 human proteins) contain very significant levels of intrinsic disorder. This is further illustrated in Figure 9, showing experimentally identified and validated interactions between the five RABV proteins, G (glycoprotein), N (nucleoprotein), L (RNA-dependent polymerase), P (phosphoprotein), and M (matrix protein), and 37 human proteins. This diagram clearly shows that most of the proteins (viral and human) in this network are "red" (highly disordered), and there are no "blue" (mostly ordered) proteins, suggesting the importance of intrinsic disorder for RABV infection.
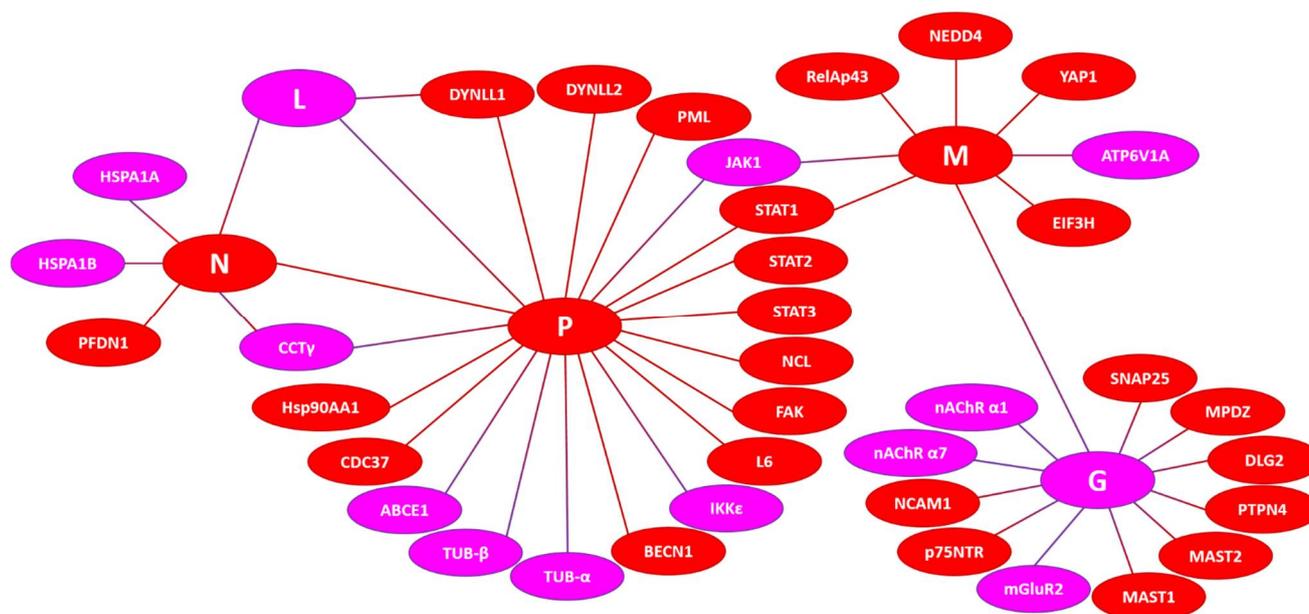

Figure 9. Disordered interactome of the RABV N-, L-, P-, M-, and G-proteins. Proteins are colored based on their PPIDR values, with highly and moderately disordered proteins being shown by red and pink colors, respectively. Note that none of the proteins in this diagram is classified as mostly ordered (there are no proteins colored in blue).

4. Conclusions

All Rabies lyssavirus PV proteins contain IDPRs, most of which are expected to aid in the flexibility of the virus and its ability to evade host antiviral defenses. All human proteins found to be RABV interactors also contain high levels of intrinsic disorder, with most of these proteins being highly disordered. This disorder-centric layer of complexity in RABV and its interaction with host proteins adds a new angle to the search for potential targets for anti-rabies drugs. Once the virus has infected a host cell, there are virtually no effective treatments available that can be used to destroy the virus. Although the modern RABV vaccine has largely eradicated the virus in developed countries, many people around

the world are unable to seek treatment until the virus has already crossed the blood–brain barrier. Without the immediate usage of the vaccine upon infection, the RABV-related mortality rate is close to 100%. The advent of bioinformatics approaches in a clinical setting has inspired many developments that have led to the creation of new drugs. It is likely that targeting regions of intrinsic disorder within the viral proteins or host proteins that directly interact with the RABV proteins will help in creating novel drugs that can target the virus once it has infected the central nervous system.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/biom12101436/s1. Figure S1: Functional disorder in human proteins interacting with the RABV P-protein. For each protein, an amino acid sequence in FASTA format is shown, followed by the disorder profile generated by RIDAO, the D^2P^2-generated functional disorder profile, the modeled 3D structure generated by AlphaFold, and the STRING-based protein–protein interaction network; Figure S2. Functional disorder in human proteins interacting with the RABV M-protein. For each protein, the amino acid sequence in FASTA format is shown, followed by the disorder profile generated by RIDAO, the D^2P^2-generated functional disorder profile, the modeled 3D structure generated by AlphaFold, and the STRING-based protein–protein interaction network; Figure S3. Functional disorder in human proteins interacting with the RABV N-protein. For each protein, the amino acid sequence in FASTA format is shown, followed by the disorder profile generated by RIDAO, the D^2P^2-generated functional disorder profile, the modeled 3D structure generated by AlphaFold, and the STRING-based protein–protein interaction network; Figure S4. Functional disorder in human proteins interacting with the RABV G-protein. For each protein, the amino acid sequence in FASTA format is shown, followed by the disorder profile generated by RIDAO, the D^2P^2-generated functional disorder profile, the modeled 3D structure generated by AlphaFold, and the STRING-based protein–protein interaction network; Figure S5. Functional disorder in human proteins interacting with the RABV L-protein. For each protein, the amino acid sequence in FASTA format is shown, followed by the disorder profile generated by RIDAO, the D^2P^2-generated functional disorder profile, the modeled 3D structure generated by AlphaFold, and the STRING-based protein–protein interaction network.

References

1. Rupprecht, C.E. Rhabdoviruses: Rabies Virus. In *Medical Microbiology*, 4th ed.; Baron, S., Ed.; University of Texas Medical Branch: Galveston, TX, USA, 1996.
2. Pieracci, E.G.; Pearson, C.M.; Wallace, R.M.; Blanton, J.D.; Whitehouse, E.R.; Ma, X.; Stauffer, K.; Chipman, R.B.; Olson, V. *Vital Signs:* Trends in Human Rabies Deaths and Exposures—United States, 1938–2018. *MMWR Morb. Mortal. Wkly. Rep.* **2019**, *68*, 524–528. [CrossRef] [PubMed]
3. Dietzschold, B.; Li, J.; Faber, M.; Schnell, M. Concepts in the pathogenesis of rabies. *Futur. Virol.* **2008**, *3*, 481–490. [CrossRef] [PubMed]
4. Gluska, S.; Zahavi, E.E.; Chein, M.; Gradus, T.; Bauer, A.; Finke, S.; Perlson, E. Rabies Virus Hijacks and Accelerates the p75NTR Retrograde Axonal Transport Machinery. *PLoS Pathog.* **2014**, *10*, e1004348. [CrossRef] [PubMed]
5. ViralZone. Lyssavirus. Available online: https://viralzone.expasy.org/resources/Rhabdoviridae_virion.jpg (accessed on 20 July 2020).
6. Okada, K.; Ito, N.; Yamaoka, S.; Masatani, T.; Ebihara, H.; Goto, H.; Nakagawa, K.; Mitake, H.; Okadera, K.; Sugiyama, M. Roles of the Rabies Virus Phosphoprotein Isoforms in Pathogenesis. *J. Virol.* **2016**, *90*, 8226–8237. [CrossRef] [PubMed]
7. Kouznetzoff, A.; Buckle, M.; Tordo, N. Identification of a region of the rabies virus N protein involved in direct binding to the viral RNA. *J. Gen. Virol.* **1998**, *79*, 1005–1013. [CrossRef]

8. Nakagawa, K.; Kobayashi, Y.; Ito, N.; Suzuki, Y.; Okada, K.; Makino, M.; Goto, H.; Takahashi, T.; Sugiyama, M. Molecular Function Analysis of Rabies Virus RNA Polymerase L Protein by Using an L Gene-Deficient Virus. *J. Virol.* **2017**, *91*, e00826-17. [CrossRef]

9. Desmézières, E.; Maillard, A.P.; Gaudin, Y.; Tordo, N.; Perrin, P. Differential stability and fusion activity of Lyssavirus glycoprotein trimers. *Virus Res.* **2002**, *91*, 181–187. [CrossRef]

10. Gupta, P.K.; Sharma, S.; Walunj, S.S.; Chaturvedi, V.K.; Raut, A.A.; Patial, S.; Rai, A.; Pandey, K.D.; Saini, M. Immunogenic and antigenic properties of recombinant soluble glycoprotein of rabies virus. *Veter. Microbiol.* **2005**, *108*, 207–214. [CrossRef]

11. Chenik, M.; Schnell, M.; Conzelmann, K.K.; Blondel, D. Mapping the Interacting Domains between the Rabies Virus Polymerase and Phosphoprotein. *J. Virol.* **1998**, *72*, 1925–1930. [CrossRef] [PubMed]

12. Pulmanausahakul, R.; Li, J.; Schnell, M.J.; Dietzschold, B. The Glycoprotein and the Matrix Protein of Rabies Virus Affect Pathogenicity by Regulating Viral Replication and Facilitating Cell-to-Cell Spread. *J. Virol.* **2008**, *82*, 2330–2338. [CrossRef]

13. Mebatsion, T.; Weiland, F.; Conzelmann, K.-K. Matrix Protein of Rabies Virus Is Responsible for the Assembly and Budding of Bullet-Shaped Particles and Interacts with the Transmembrane Spike Glycoprotein G. *J. Virol.* **1999**, *73*, 242–250. [CrossRef] [PubMed]

14. Chenik, M.; Chebli, K.; Blondel, D. Translation initiation at alternate in-frame AUG codons in the rabies virus phosphoprotein mRNA is mediated by a ribosomal leaky scanning mechanism. *J. Virol.* **1995**, *69*, 707–712. [CrossRef] [PubMed]

15. Li, Y.; Dong, W.; Shi, Y.; Deng, F.; Chen, X.; Wan, C.; Zhou, M.; Zhao, L.; Fu, Z.F.; Peng, G. Rabies virus phosphoprotein interacts with ribosomal protein L9 and affects rabies virus replication. *Virology* **2015**, *488*, 216–224. [CrossRef] [PubMed]

16. Wiltzer, L.; Okada, K.; Yamaoka, S.; Larrous, F.; Kuusisto, H.V.; Sugiyama, M.; Blondel, D.; Bourhy, H.; Jans, D.; Ito, N.; et al. Interaction of Rabies Virus P-Protein with STAT Proteins is Critical to Lethal Rabies Disease. *J. Infect. Dis.* **2013**, *209*, 1744–1753. [CrossRef] [PubMed]

17. Vidy, A.; El Bougrini, J.; Chelbi-Alix, M.K.; Blondel, D. The Nucleocytoplasmic Rabies Virus P Protein Counteracts Interferon Signaling by Inhibiting both Nuclear Accumulation and DNA Binding of STAT1. *J. Virol.* **2007**, *81*, 4255–4263. [CrossRef] [PubMed]

18. Moseley, G.W.; Lahaye, X.; Roth, D.M.; Oksayan, S.; Filmer, R.P.; Rowe, C.L.; Blondel, D.; Jans, D. Dual modes of rabies P-protein association with microtubules: A novel strategy to suppress the antiviral response. *J. Cell Sci.* **2009**, *122*, 3652–3662. [CrossRef] [PubMed]

19. Masatani, T.; Ito, N.; Ito, Y.; Nakagawa, K.; Abe, M.; Yamaoka, S.; Okadera, K.; Sugiyama, M. Importance of rabies virus nucleoprotein in viral evasion of interferon response in the brain. *Microbiol. Immunol.* **2013**, *57*, 511–517. [CrossRef] [PubMed]

20. Masatani, T.; Ito, N.; Shimizu, K.; Ito, Y.; Nakagawa, K.; Sawaki, Y.; Koyama, H.; Sugiyama, M. Rabies Virus Nucleoprotein Functions to Evade Activation of the RIG-I-Mediated Antiviral Response. *J. Virol.* **2010**, *84*, 4002–4012. [CrossRef] [PubMed]

21. Ogino, M.; Ito, N.; Sugiyama, M.; Ogino, T. The Rabies Virus L Protein Catalyzes mRNA Capping with GDP Polyribonucleotidyl-transferase Activity. *Viruses* **2016**, *8*, 144. [CrossRef] [PubMed]

22. Finke, S.; Mueller-Waldeck, R.; Conzelmann, K.-K. Rabies virus matrix protein regulates the balance of virus transcription and replication. *J. Gen. Virol.* **2003**, *84*, 1613–1621. [CrossRef]

23. Ben Khalifa, Y.; Luco, S.; Besson, B.; Sonthonnax, F.; Archambaud, M.; Grimes, J.M.; Larrous, F.; Bourhy, H. The matrix protein of rabies virus binds to RelAp43 to modulate NF-κB-dependent gene expression related to innate immunity. *Sci. Rep.* **2016**, *6*, 39420. [CrossRef] [PubMed]

24. Zhang, G.; Wang, H.; Mahmood, F.; Fu, Z.F. Rabies virus glycoprotein is an important determinant for the induction of innate immune responses and the pathogenic mechanisms. *Vet. Microbiol.* **2013**, *162*, 601–613. [CrossRef] [PubMed]

25. Dunker, A.K.; Lawson, J.D.; Brown, C.J.; Williams, R.M.; Romero, P.; Oh, J.S.; Oldfield, C.J.; Campen, A.M.; Ratliff, C.M.; Hipps, K.W.; et al. Intrinsically disordered protein. *J. Mol. Graph. Model* **2001**, *19*, 26–59. [CrossRef]

26. Dunker, A.K.; Brown, C.J.; Lawson, J.D.; Iakoucheva, L.M.; Obradovic, Z. Intrinsic disorder and protein function. *Biochemistry* **2002**, *41*, 6573–6582. [CrossRef] [PubMed]

27. Tompa, P. Intrinsically unstructured proteins. *Trends Biochem. Sci.* **2002**, *27*, 527–533. [CrossRef]

28. Uversky, V.N. Natively unfolded proteins: A point where biology waits for physics. *Protein Sci.* **2002**, *11*, 739–756. [CrossRef]

29. Uversky, V.N. What does it mean to be natively unfolded? *Eur. J. Biochem.* **2002**, *269*, 2–12. [CrossRef]

30. Uversky, V.N.; Dunker, A.K. Understanding protein non-folding. *Biochim. Biophys. Acta BBA Proteins Proteom.* **2010**, *1804*, 1231–1264. [CrossRef]

31. Uversky, V.N.; Gillespie, J.R.; Fink, A.L. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* **2000**, *41*, 415–427. [CrossRef]

32. Dunker, A.K.; Cortese, M.; Romero, P.; Iakoucheva, L.; Uversky, V.N. Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J.* **2005**, *272*, 5129–5148. [CrossRef]

33. Uversky, V.N. Protein intrinsic disorder and structure-function continuum. *Prog. Mol. Biol. Transl. Sci.* **2019**, *166*, 1–17. [CrossRef] [PubMed]

34. Dunker, A.K.; Obradovic, Z.; Romero, P.; Garner, E.C.; Brown, C.J. Intrinsic protein disorder in complete genomes. *Genome Inform.* **2000**, *11*, 161–171.

35. Ward, J.J.; Sodhi, J.S.; McGuffin, L.J.; Buxton, B.F.; Jones, D.T. Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. *J. Mol. Biol.* **2004**, *337*, 635–645. [CrossRef] [PubMed]

36. Xue, B.; Dunker, A.K.; Uversky, V.N. Orderly order in protein intrinsic disorder distribution: Disorder in 3500 proteomes from viruses and the three domains of life. *J. Biomol. Struct. Dyn.* **2012**, *30*, 137–149. [CrossRef]
37. Peng, Z.; Yan, J.; Fan, X.; Mizianty, M.J.; Xue, B.; Wang, K.; Hu, G.; Uversky, V.N.; Kurgan, L. Exceptionally abundant exceptions: Comprehensive characterization of intrinsic disorder in all domains of life. *Cell Mol. Life Sci.* **2014**, *72*, 137–151. [CrossRef]
38. Uversky, V.N. The mysterious unfoldome: Structureless, underappreciated, yet vital part of any given proteome. *J. Biomed. Biotechnol.* **2010**, *2010*, 568068. [CrossRef]
39. Iakoucheva, L.M.; Brown, C.J.; Lawson, J.D.; Obradovic, Z.; Dunker, A.K. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J. Mol. Biol.* **2002**, *323*, 573–584. [CrossRef]
40. Uversky, V.N. Unusual biophysics of intrinsically disordered proteins. *Biochim. Biophys. Acta BBA Proteins Proteom.* **2012**, *1834*, 932–951. [CrossRef]
41. Dyson, H.J.; Wright, P. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197–208. [CrossRef]
42. E Wright, P.; Dyson, H.J. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **1999**, *293*, 321–331. [CrossRef]
43. Tompa, P. The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett.* **2005**, *579*, 3346–3354. [CrossRef] [PubMed]
44. Radivojac, P.; Iakoucheva, L.M.; Oldfield, C.J.; Obradovic, Z.; Uversky, V.N.; Dunker, A.K. Intrinsic Disorder and Functional Proteomics. *Biophys. J.* **2007**, *92*, 1439–1456. [CrossRef]
45. Vucetic, S.; Xie, H.; Iakoucheva, L.M.; Oldfield, C.J.; Dunker, A.K.; Obradovic, Z.; Uversky, V.N. Functional Anthology of Intrinsic Disorder. 2. Cellular Components, Domains, Technical Terms, Developmental Processes, and Coding Sequence Diversities Correlated with Long Disordered Regions. *J. Proteome Res.* **2007**, *6*, 1899–1916. [CrossRef] [PubMed]
46. Xie, H.; Vucetic, S.; Iakoucheva, L.M.; Oldfield, C.J.; Dunker, A.K.; Obradovic, Z.; Uversky, V.N. Functional Anthology of Intrinsic Disorder. 3. Ligands, Post-Translational Modifications, and Diseases Associated with Intrinsically Disordered Proteins. *J. Proteome Res.* **2007**, *6*, 1917–1932. [CrossRef]
47. Xie, H.; Vucetic, S.; Iakoucheva, L.M.; Oldfield, C.J.; Dunker, A.K.; Uversky, V.N.; Obradovic, Z. Functional Anthology of Intrinsic Disorder. 1. Biological Processes and Functions of Proteins with Long Disordered Regions. *J. Proteome Res.* **2007**, *6*, 1882–1898. [CrossRef] [PubMed]
48. Uversky, V.N.; Oldfield, C.J.; Dunker, A.K. Intrinsically Disordered Proteins in Human Diseases: Introducing the D^2 Concept. *Annu. Rev. Biophys.* **2008**, *37*, 215–246. [CrossRef] [PubMed]
49. Vacic, V.; Markwick, P.R.L.; Oldfield, C.J.; Zhao, X.; Haynes, C.; Uversky, V.N.; Iakoucheva, L.M. Disease-Associated Mutations Disrupt Functionally Important Regions of Intrinsic Protein Disorder. *PLoS Comput. Biol.* **2012**, *8*, e1002709. [CrossRef] [PubMed]
50. Dunker, A.K.; Garner, E.; Guilliot, S.; Romero, P.; Albrecht, K.; Hart, J.; Obradovic, Z.; Kissinger, C.; E Villafranca, J. Protein disorder and the evolution of molecular recognition: Theory, predictions and observations. *Pac. Symp. Biocomput. Pac. Symp. Biocomput.* **1998**, *3*, 473–484.
51. Daughdrill, G.W.; Pielak, G.J.; Uversky, V.N.; Cortese, M.S.; Dunker, A.K. Natively disordered proteins. In *Handbook of Protein Folding*; Buchner, J., Kiefhaber, T., Eds.; Wiley-VCH, Verlag GmbH & Co. KGaA: Weinheim, Germany, 2005; pp. 271–353.
52. Uversky, V.N. Intrinsic Disorder-based Protein Interactions and their Modulators. *Curr. Pharm. Des.* **2013**, *19*, 4191–4213. [CrossRef]
53. Uversky, V.N. Functional roles of transiently and intrinsically disordered regions within proteins. *FEBS J.* **2015**, *282*, 1182–1189. [CrossRef]
54. Uversky, V.N. p53 Proteoforms and Intrinsic Disorder: An Illustration of the Protein Structure–Function Continuum Concept. *Int. J. Mol. Sci.* **2016**, *17*, 1874. [CrossRef] [PubMed]
55. Smith, L.M.; The Consortium for Top Down Proteomics; Kelleher, N.L. Proteoform: A single term describing protein complexity. *Nat. Methods* **2013**, *10*, 186–187. [CrossRef] [PubMed]
56. Xue, B.; Ganti, K.; Rabionet, A.; Banks, L.; Uversky, V. Disordered Interactome of Human Papillomavirus. *Curr. Pharm. Des.* **2014**, *20*, 1274–1292. [CrossRef] [PubMed]
57. Uversky, V.N.; Roman, A.; Oldfield, A.C.J.; Dunker, A.K. Protein Intrinsic Disorder and Human Papillomaviruses: Increased Amount of Disorder in E6 and E7 Oncoproteins from High Risk HPVs. *J. Proteome Res.* **2006**, *5*, 1829–1842. [CrossRef]
58. Xue, B.; Mizianty, M.J.; Kurgan, L.; Uversky, V.N. Protein intrinsic disorder as a flexible armor and a weapon of HIV-1. *Experientia* **2011**, *69*, 1211–1259. [CrossRef] [PubMed]
59. Dolan, P.; Roth, A.P.; Xue, B.; Sun, R.; Dunker, A.K.; Uversky, V.N.; LaCount, D.J. Intrinsic disorder mediates hepatitis C virus core-host cell protein interactions. *Protein Sci.* **2014**, *24*, 221–235. [CrossRef] [PubMed]
60. Fan, X.; Xue, B.; Dolan, P.T.; LaCount, D.J.; Kurgan, L.; Uversky, V.N. The intrinsic disorder status of the human hepatitis C virus proteome. *Mol. BioSyst.* **2014**, *10*, 1345–1363. [CrossRef]
61. Meng, F.; Badierah, R.A.; Almehdar, H.A.; Redwan, E.M.; Kurgan, L.; Uversky, V.N. Unstructural biology of the dengue virus proteins. *FEBS J.* **2015**, *282*, 3368–3394. [CrossRef]
62. Kumar, D.; Singh, A.; Kumar, P.; Uversky, V.N.; Rao, C.D.; Giri, R. Understanding the penetrance of intrinsic protein disorder in rotavirus proteome. *Int. J. Biol. Macromol.* **2019**, *144*, 892–908. [CrossRef]
63. Whelan, J.N.; Reddy, K.D.; Uversky, V.N.; Teng, M.N. Functional correlations of respiratory syncytial virus proteins to intrinsic disorder. *Mol. BioSyst.* **2016**, *12*, 1507–1526. [CrossRef]

64. Mishra, P.M.; Uversky, V.N.; Giri, R. Molecular Recognition Features in Zika Virus Proteome. *J. Mol. Biol.* **2017**, *430*, 2372–2388. [CrossRef] [PubMed]

65. Giri, R.; Kumar, D.; Sharma, N.; Uversky, V.N. Intrinsically Disordered Side of the Zika Virus Proteome. *Front. Cell. Infect. Microbiol.* **2016**, *6*, 144. [CrossRef] [PubMed]

66. Singh, A.; Kumar, A.; Yadav, R.; Uversky, V.N.; Giri, R. Deciphering the dark proteome of Chikungunya virus. *Sci. Rep.* **2018**, *8*, 5822. [CrossRef] [PubMed]

67. Redwan, E.M.; AlJaddawi, A.A.; Uversky, V.N. Structural disorder in the proteome and interactome of Alkhurma virus (ALKV). *Experientia* **2018**, *76*, 577–608. [CrossRef] [PubMed]

68. Bhardwaj, T.; Saumya, K.U.; Kumar, P.; Sharma, N.; Gadhave, K.; Uversky, V.N.; Giri, R. Japanese encephalitis virus—Exploring the dark proteome and disorder-function paradigm. *FEBS J.* **2020**, *287*, 3751–3776. [CrossRef] [PubMed]

69. Giri, R.; Bhardwaj, T.; Shegane, M.; Gehi, B.R.; Kumar, P.; Gadhave, K.; Oldfield, C.J.; Uversky, V.N. Understanding the COVID-19 via Comparative Analysis of Dark Proteomes of SARS-CoV-2, Human SARS and Bat SARS-Like Coronaviruses. *Cell Mol. Life Sci.* **2021**, *78*, 1655–1688. [CrossRef]

70. UniProt Consortium. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **2019**, *47*, D506–D515. [CrossRef]

71. Zandi, F.; Goshadrou, F.; Meyfour, A.; Vaziri, B. Rabies Infection: An Overview of Lyssavirus-Host Protein Interactions. *Iran. Biomed. J.* **2021**, *25*, 226–242. [CrossRef]

72. Chen, J.; Kriwacki, R.W. Intrinsically Disordered Proteins: Structure, Function and Therapeutics. *J. Mol. Biol.* **2018**, *430*, 2275–2277. [CrossRef] [PubMed]

73. Xue, B.; Dunbrack, R.L.; Williams, R.W.; Dunker, A.K.; Uversky, V.N. PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *Biochim. Biophys. Acta BBA Proteins Proteom.* **2010**, *1804*, 996–1010. [CrossRef]

74. Dosztanyi, Z.; Csizmók, V.; Tompa, P.; Simon, I. The Pairwise Energy Content Estimated from Amino Acid Composition Discriminates between Folded and Intrinsically Unstructured Proteins. *J. Mol. Biol.* **2005**, *347*, 827–839. [CrossRef] [PubMed]

75. Mészáros, B.; Erdős, G.; Dosztányi, Z. IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **2018**, *46*, W329–W337. [CrossRef] [PubMed]

76. Mészáros, B.; Simon, I.; Dosztanyi, Z. Prediction of Protein Binding Regions in Disordered Proteins. *PLoS Comput. Biol.* **2009**, *5*, e1000376. [CrossRef] [PubMed]

77. Dosztanyi, Z.; Meszaros, B.; Simon, I. ANCHOR: Web server for predicting protein binding regions in disordered proteins. *Bioinformatics* **2009**, *25*, 2745–2746. [CrossRef] [PubMed]

78. Dayhoff, G.W.I.; Uversky, V.N. Rapid prediction and analysis of protein intrinsic disorder. *Protein Sci.* 2022; *in press*.

79. Romero, P.; Obradovic, Z.; Li, X.; Garner, E.C.; Brown, C.J.; Dunker, A.K. Sequence complexity of disordered protein. *Proteins* **2001**, *42*, 38–48. [CrossRef]

80. Peng, K.; Radivojac, P.; Vucetic, S.; Dunker, A.K.; Obradovic, Z. Length-dependent prediction of protein intrinsic disorder. *BMC Bioinform.* **2006**, *7*, 208. [CrossRef] [PubMed]

81. Peng, K.; Vucetic, S.; Radivojac, P.; Brown, C.J.; Dunker, A.K.; Obradovic, Z. Optimizing long intrinsic disorder predictors with protein evolutionary information. *J. Bioinform. Comput. Biol.* **2005**, *3*, 35–60. [CrossRef] [PubMed]

82. Dosztanyi, Z.; Csizmok, V.; Tompa, P.; Simon, I. IUPred: Web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **2005**, *21*, 3433–3434. [CrossRef] [PubMed]

83. Mohan, A.; Sullivan, W.J., Jr.; Radivojac, P.; Dunker, A.K.; Uversky, V.N. Intrinsic disorder in pathogenic and non-pathogenic microbes: Discovering and analyzing the unfoldomes of early-branching eukaryotes. *Mol. BioSyst.* **2008**, *4*, 328–340. [CrossRef] [PubMed]

84. Sun, X.; Xue, B.; Jones, W.T.; Rikkerink, E.; Dunker, A.K.; Uversky, V.N. A functionally required unfoldome from the plant kingdom: Intrinsically disordered N-terminal domains of GRAS proteins are involved in molecular recognition during plant development. *Plant Mol. Biol.* **2011**, *77*, 205–223. [CrossRef]

85. Xue, B.; Oldfield, C.J.; Van, Y.-Y.; Dunker, A.K.; Uversky, V.N. Protein intrinsic disorder and induced pluripotent stem cells. *Mol. BioSyst.* **2011**, *8*, 134–150. [CrossRef] [PubMed]

86. Huang, F.; Oldfield, C.; Meng, J.; Hsu, W.-L.; Xue, B.; Uversky, V.N.; Romero, P.; Dunker, A.K. Subclassifying disordered proteins by the CH-CDF plot method. *Biocomputing* **2011**, 128–139. [CrossRef]

87. Rajagopalan, K.; Mooney, S.M.; Parekh, N.; Getzenberg, R.H.; Kulkarni, P. A majority of the cancer/testis antigens are intrinsically disordered proteins. *J. Cell. Biochem.* **2011**, *112*, 3256–3267. [CrossRef] [PubMed]

88. Oates, M.E.; Romero, P.; Ishida, T.; Ghalwash, M.; Mizianty, M.J.; Xue, B.; Dosztányi, Z.; Uversky, V.N.; Obradovic, Z.; Kurgan, L.; et al. D^{2}P^2: Database of disordered protein predictions. *Nucleic Acids Res.* **2013**, *41*, D508–D516. [CrossRef] [PubMed]

89. Ishida, T.; Kinoshita, K. PrDOS: Prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* **2007**, *35*, W460–W464. [CrossRef]

90. Walsh, I.; Martin, A.J.M.; Di Domenico, T.; Tosatto, S.C.E. ESpritz: Accurate and fast prediction of protein disorder. *Bioinformatics* **2012**, *28*, 503–509. [CrossRef] [PubMed]

91. Andreeva, A.; Howorth, D.; Brenner, S.E.; Hubbard, T.J.; Chothia, C.; Murzin, A.G. SCOP database in 2004: Refinements integrate structure and sequence family data. *Nucleic Acids Res.* **2004**, *32* (Suppl. S1), D226–D229. [CrossRef]

92. Murzin, A.G.; Brenner, S.E.; Hubbard, T.; Chothia, C. SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **1995**, *247*, 536–540. [CrossRef]

93.	De Lima Morais, D.A.; Fang, H.; Rackham, O.J.L.; Wilson, D.; Pethica, R.; Chothia, C.; Gough, J. SUPERFAMILY 1.75 including a domain-centric gene ontology method. *Nucleic Acids Res.* **2011**, *39*, D427–D434. [CrossRef] [PubMed]

94.	Hornbeck, P.V.; Kornhauser, J.M.; Tkachev, S.; Zhang, B.; Skrzypek, E.; Murray, B.; Latham, V.; Sullivan, M. PhosphoSitePlus: A comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res.* **2011**, *40*, D261–D270. [CrossRef]

95.	Szklarczyk, D.; Franceschini, A.; Kuhn, M.; Simonovic, M.; Roth, A.; Minguez, P.; Doerks, T.; Stark, M.; Muller, J.; Bork, P.; et al. The STRING database in 2011: Functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* **2010**, *39*, D561–D568. [CrossRef] [PubMed]

96.	Hardenberg, M.; Horvath, A.; Ambrus, V.; Fuxreiter, M.; Vendruscolo, M. Widespread occurrence of the droplet state of proteins in the human proteome. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 33254–33262. [CrossRef] [PubMed]

97.	Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; De Beer, T.A.P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, W296–W303. [CrossRef]

98.	Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [CrossRef]

99.	Nevers, Q.; Albertini, A.A.; Lagaudrière-Gesbert, C.; Gaudin, Y. Negri bodies and other virus membrane-less replication compartments. *Biochim. Biophys. Acta* **2020**, *1867*, 118831. [CrossRef] [PubMed]

100.	Nikolic, J.; Le Bars, R.; Lama, Z.; Scrima, N.; Lagaudrière-Gesbert, C.; Gaudin, Y.; Blondel, D. Negri bodies are viral factories with properties of liquid organelles. *Nat. Commun.* **2017**, *8*, 58. [CrossRef]

101.	Brzózka, K.; Finke, S.; Conzelmann, K.-K. Identification of the Rabies Virus Alpha/Beta Interferon Antagonist: Phosphoprotein P Interferes with Phosphorylation of Interferon Regulatory Factor 3. *J. Virol.* **2005**, *79*, 7673–7681. [CrossRef]

102.	Vidy, A.; Chelbi-Alix, M.; Blondel, D. Rabies Virus P Protein Interacts with STAT1 and Inhibits Interferon Signal Transduction Pathways. *J. Virol.* **2005**, *79*, 14411–14420. [CrossRef] [PubMed]

103.	Lieu, K.G.; Brice, A.; Wiltzer, L.; Hirst, B.; Jans, D.A.; Blondel, D.; Moseley, G.W. The Rabies Virus Interferon Antagonist P Protein Interacts with Activated STAT3 and Inhibits Gp130 Receptor Signaling. *J. Virol.* **2013**, *87*, 8261–8265. [CrossRef] [PubMed]

104.	Niu, X.; Tang, L.; Tseggai, T.; Guo, Y.; Fu, Z.F. Wild-type rabies virus phosphoprotein is associated with viral sensitivity to type I interferon treatment. *Arch. Virol.* **2013**, *158*, 2297–2305. [CrossRef]

105.	Gupta, A.K.; Blondel, D.; Choudhary, S.; Banerjee, A.K. The Phosphoprotein of Rabies Virus Is Phosphorylated by a Unique Cellular Protein Kinase and Specific Isomers of Protein Kinase C. *J. Virol.* **2000**, *74*, 91–98. [CrossRef] [PubMed]

106.	Zhan, J.; Watts, E.; Brice, A.M.; Metcalfe, R.D.; Rozario, A.M.; Sethi, A.; Yan, F.; Bell, T.D.M.; Griffin, M.D.W.; Moseley, G.W.; et al. Molecular Basis of Functional Effects of Phosphorylation of the C-Terminal Domain of the Rabies Virus P Protein. *J. Virol.* **2022**, *96*, e0011122. [CrossRef] [PubMed]

107.	Brice, A.; Whelan, D.; Ito, N.; Shimizu, K.; Wiltzer-Bach, L.; Lo, C.; Blondel, D.; Jans, D.; Bell, T.D.M.; Moseley, G.W. Quantitative Analysis of the Microtubule Interaction of Rabies Virus P3 Protein: Roles in Immune Evasion and Pathogenesis. *Sci. Rep.* **2016**, *6*, 33493. [CrossRef]

108.	Rowe, C.L.; Wagstaff, K.; Oksayan, S.; Glover, D.J.; Jans, D.; Moseley, G.W. Nuclear Trafficking of the Rabies Virus Interferon Antagonist P-Protein Is Regulated by an Importin-Binding Nuclear Localization Sequence in the C-Terminal Domain. *PLoS ONE* **2016**, *11*, e0150477. [CrossRef] [PubMed]

109.	Pasdeloup, D.; Poisson, N.; Raux, H.; Gaudin, Y.; Ruigrok, R.W.; Blondel, D. Nucleocytoplasmic shuttling of the rabies virus P protein requires a nuclear localization signal and a CRM1-dependent nuclear export signal. *Virology* **2005**, *334*, 284–293. [CrossRef] [PubMed]

110.	Blondel, D.; Regad, T.; Poisson, N.; Pavie, B.; Harper, F.; Pandolfi, P.P.; de Thé, H.; Chelbi-Alix, M.K. Rabies virus P and small P products interact directly with PML and reorganize PML nuclear bodies. *Oncogene* **2002**, *21*, 7957–7970. [CrossRef]

111.	Horvath, A.; Miskei, M.; Ambrus, V.; Vendruscolo, M.; Fuxreiter, M. Sequence-based prediction of protein binding mode landscapes. *PLoS Comput. Biol.* **2020**, *16*, e1007864. [CrossRef]

112.	Goh, G.K.-M.; Dunker, A.K.; Foster, J.A.; Uversky, V.N. Shell disorder analysis predicts greater resilience of the SARS-CoV-2 (COVID-19) outside the body and in body fluids. *Microb. Pathog.* **2020**, *144*, 104177. [CrossRef] [PubMed]

113.	Goh, G.K.-M.; Dunker, A.K.; Foster, J.A.; Uversky, V.N. Rigidity of the Outer Shell Predicted by a Protein Intrinsic Disorder Model Sheds Light on the COVID-19 (Wuhan-2019-nCoV) Infectivity. *Biomolecules* **2020**, *10*, 331. [CrossRef]

114.	Goh, G.K.-M.; Dunker, A.K.; Foster, J.A.; Uversky, V.N. Nipah shell disorder, modes of infection, and virulence. *Microb. Pathog.* **2020**, *141*, 103976. [CrossRef]

115.	Goh, G.K.M.; Dunker, A.K.; Foster, J.A.; Uversky, V.N. Zika and Flavivirus Shell Disorder: Virulence and Fetal Morbidity. *Biomolecules* **2019**, *9*, 710. [CrossRef] [PubMed]

116.	Goh, G.K.M.; Dunker, A.K.; Foster, J.A.; Uversky, V.N. HIV Vaccine Mystery and Viral Shell Disorder. *Biomolecules* **2019**, *9*, 178. [CrossRef]

117.	Goh, G.K.-M.; Dunker, A.K.; Uversky, V.N. Correlating Flavivirus virulence and levels of intrinsic disorder in shell proteins: Protective roles vs. immune evasion. *Mol. BioSyst.* **2016**, *12*, 1881–1891. [CrossRef]

118.	Goh, G.K.-M.; Dunker, A.K.; Uversky, V.N. Shell disorder, immune evasion and transmission behaviors among human and animal retroviruses. *Mol. BioSyst.* **2015**, *11*, 2312–2323. [CrossRef]

119. Goh, G.K.-M.; Dunker, A.K.; Uversky, V. Prediction of Intrinsic Disorder in MERS-CoV/HCoV-EMC Supports a High Oral-Fecal Transmission. *PLoS Curr.* **2013**, *5*, 24270586. [CrossRef]

120. Goh, G.K.-M.; Dunker, A.K.; Uversky, V.N. Understanding Viral Transmission Behavior via Protein Intrinsic Disorder Prediction: Coronaviruses. *J. Pathog.* **2012**, *2012*, 738590. [CrossRef] [PubMed]

121. Montespan, C.; Wiethoff, C.M.; Wodrich, H. A Small Viral PPxY Peptide Motif to Control Antiviral Autophagy. *J. Virol.* **2017**, *91*, e00581-17. [CrossRef]

122. Graham, S.; Assenberg, R.; Delmas, O.; Verma, A.; Gholami, A.; Talbi, C.; Owens, R.; Stuart, D.; Grimes, J.M.; Bourhy, H. Rhabdovirus Matrix Protein Structures Reveal a Novel Mode of Self-Association. *PLoS Pathog.* **2008**, *4*, e1000251. [CrossRef] [PubMed]

123. Realegeno, S.; Niezgoda, M.; Yager, P.A.; Kumar, A.; Hoque, L.; Orciari, L.; Sambhara, S.; Olson, V.A.; Satheshkumar, P.S. An ELISA-based method for detection of rabies virus nucleoprotein-specific antibodies in human antemortem samples. *PLoS ONE* **2018**, *13*, e0207009. [CrossRef]

124. Lingappa, U.F.; Wu, X.; Macieik, A.; Yu, S.F.; Atuegbu, A.; Corpuz, M.; Francis, J.; Nichols, C.; Calayag, A.; Shi, H.; et al. Host–rabies virus protein–protein interactions as druggable antiviral targets. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, E861–E868. [CrossRef] [PubMed]

125. Mavrakis, M.; Iseni, F.; Mazzaa, C.; Schoehnac, G.; Ebelc, C.; Gentzeld, M.; Franzd, T.; Ruigrok, R.W. Isolation and Characterisation of the Rabies Virus N°-P Complex Produced in Insect Cells. *Virology* **2003**, *305*, 406–414. [CrossRef] [PubMed]

126. Albertini, A.A.V.; Wernimont, A.K.; Muziol, T.M.; Ravelli, R.B.G.; Clapier, C.R.; Schoehn, G.; Weissenhorn, W.; Ruigrok, R.W.H. Crystal Structure of the Rabies Virus Nucleoprotein-RNA Complex. *Science* **2006**, *313*, 360–363. [CrossRef] [PubMed]

127. Luo, M.; Green, T.J.; Zhang, X.; Tsao, J.; Qiu, S. Conserved characteristics of the rhabdovirus nucleoprotein. *Virus Res.* **2007**, *129*, 246–251. [CrossRef] [PubMed]

128. Etessami, R.; Conzelmann, K.-K.; Fadai-Ghotbi, B.; Natelson, B.; Tsiang, H.; Ceccaldi, P.-E. Spread and pathogenic characteristics of a G-deficient rabies virus recombinant: An in vitro and in vivo study. *J. Gen. Virol.* **2000**, *81*, 2147–2153. [CrossRef] [PubMed]

129. Nitschel, S.; Zaeck, L.M.; Potratz, M.; Nolden, T.; Kamp, V.T.; Franzke, K.; Höper, D.; Pfaff, F.; Finke, S. Point Mutations in the Glycoprotein Ectodomain of Field Rabies Viruses Mediate Cell Culture Adaptation through Improved Virus Release in a Host Cell Dependent and Independent Manner. *Viruses* **2021**, *13*, 1989. [CrossRef]

130. Yang, F.; Lin, S.; Ye, F.; Yang, J.; Qi, J.; Chen, Z.; Lin, X.; Wang, J.; Yue, D.; Cheng, Y.; et al. Structural Analysis of Rabies Virus Glycoprotein Reveals pH-Dependent Conformational Changes and Interactions with a Neutralizing Antibody. *Cell Host Microbe* **2020**, *27*, 441–453.e7. [CrossRef]

131. Schnell, M.J.; Conzelmann, K.-K. Polymerase Activity ofin VitroMutated Rabies Virus L Protein. *Virology* **1995**, *214*, 522–530. [CrossRef]

132. Horwitz, J.A.; Jenni, S.; Harrison, S.C.; Whelan, S.P.J. Structure of a rabies virus polymerase complex from electron cryo-microscopy. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 2099–2107. [CrossRef]

133. Liu, X.; Nawaz, Z.; Guo, C.; Ali, S.; Naeem, M.A.; Jamil, T.; Ahmad, W.; Siddiq, M.U.; Ahmed, S.; Idrees, M.A.; et al. Rabies Virus Exploits Cytoskeleton Network to Cause Early Disease Progression and Cellular Dysfunction. *Front. Vet. Sci.* **2022**, *9*, 889873. [CrossRef]

134. Kammouni, W.; Wood, H.; Saleh, A.; Appolinario, C.M.; Fernyhough, P.; Jackson, A.C. Rabies virus phosphoprotein interacts with mitochondrial Complex I and induces mitochondrial dysfunction and oxidative stress. *J. Neurovirol.* **2015**, *21*, 370–382. [CrossRef]

135. Sharma, L.K.; Lu, J.; Bai, Y. Mitochondrial Respiratory Complex I: Structure, Function and Implication in Human Diseases. *Curr. Med. Chem.* **2009**, *16*, 1266–1277. [CrossRef] [PubMed]

136. Wang, J.; Wang, Z.; Liu, R.; Shuai, L.; Wang, X.; Luo, J.; Wang, C.; Chen, W.; Wang, X.; Ge, J.; et al. Metabotropic glutamate receptor subtype 2 is a cellular receptor for rabies virus. *PLoS Pathog.* **2018**, *14*, e1007189. [CrossRef] [PubMed]

137. Tuffereau, C.; Bénéjean, J.; Blondel, D.; Kieffer, B.; Flamand, A. Low-affinity nerve-growth factor receptor (P75NTR) can serve as a receptor for rabies virus. *EMBO J.* **1998**, *17*, 7250–7259. [CrossRef]

138. Thoulouze, M.-I.; Lafage, M.; Schachner, M.; Hartmann, U.; Cremer, H.; Lafon, M. The Neural Cell Adhesion Molecule Is a Receptor for Rabies Virus. *J. Virol.* **1998**, *72*, 7181–7190. [CrossRef] [PubMed]

139. Lentz, T.L.; Burrage, T.G.; Smith, A.L.; Crick, J.; Tignor, G.H. Is the Acetylcholine Receptor a Rabies Virus Receptor? *Science* **1982**, *215*, 182–184. [CrossRef] [PubMed]

140. Sajjanar, B.; Dhusia, K.; Saxena, S.; Joshi, V.; Bisht, D.; Thakuria, D.; Manjunathareddy, G.B.; Ramteke, P.W.; Kumar, S. Nicotinic acetylcholine receptor alpha 1(nAChRα1) subunit peptides as potential antiviral agents against rabies virus. *Int. J. Biol. Macromol.* **2017**, *104*, 180–188. [CrossRef] [PubMed]

141. Embregts, C.W.E.; Begeman, L.; Voesenek, C.J.; Martina, B.E.E.; Koopmans, M.P.G.; Kuiken, T.; GeurtsvanKessel, C.H. Street RABV Induces the Cholinergic Anti-inflammatory Pathway in Human Monocyte-Derived Macrophages by Binding to nAChr α7. *Front. Immunol.* **2021**, *12*, 622516. [CrossRef] [PubMed]

142. Peng, Z.L.; Kurgan, L. Comprehensive Comparative Assessment of In-Silico Predictors of Disordered Regions. *Curr. Protein Pept. Sci.* **2012**, *13*, 6–18. [CrossRef] [PubMed]

143. Necci, M.; Piovesan, D.; CAID Predictors; Curators, D.; Tosatto, S.C.E. Critical assessment of protein intrinsic disorder prediction. *Nat. Methods* **2021**, *18*, 472–481. [CrossRef]